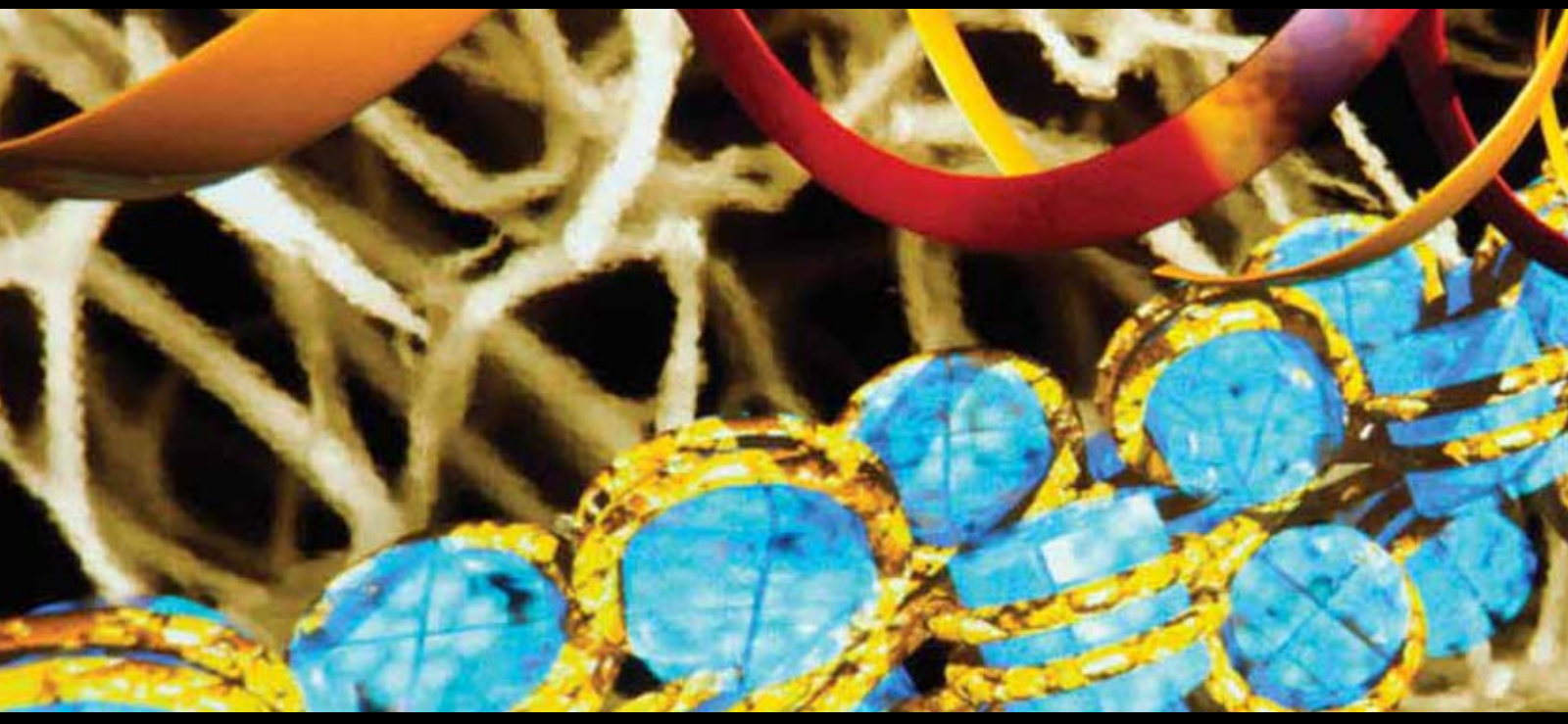
Molecular Biology International

DNA in 3R: Repair, Replication, and Recombination

Guest Editors: Frédéric Coin, Bernardo Reina-San-Martin, Giuseppina Giglia-Mari, and Mark Berneburg



DNA in 3R: Repair, Replication, and Recombination

Molecular Biology International

DNA in 3R: Repair, Replication, and Recombination

Guest Editors: Frédéric Coin, Bernardo Reina-San-Martin, Giuseppina Giglia-Mari, and Mark Berneburg

Editorial Board

Contents

Hindawi Publishing Corporation
Molecular Biology International
Volume 2012, Article ID 658579, 1 page
doi:10.1155/2012/658579

Editorial

DNA in 3R: Repair, Replication, and Recombination

Frédéric Coin,[1] Bernardo Reina-San-Martin,[1] Giuseppina Giglia-Mari,[2] and Mark Berneburg[3]

[1] *IGBMC, 67404 Illkirch Cedex, France*
[2] *IPBS, 31077 Toulouse, France*
[3] *University of Tuebingen, Germany*

Correspondence should be addressed to Frédéric Coin, fredr@igbmc.fr

Received 20 December 2011; Accepted 20 December 2011

This special issue entitled "DNA in 3R: Repair, Replication, and Recombination" is dedicated to biological processes that preserve the integrity of our genome. These phenomena have attracted broad interest among a large community of scientists that cross disciplines from mathematics, physics, chemistry, and biology to clinical scientists. DNA is continuously exposed to a range of damaging agents, including reactive cellular metabolites, environmental chemicals, ionizing radiation, and UV light. The biochemical consequences of DNA lesions are diverse and range from obstruction of fundamental cellular pathways like transcription and replication to fixation of mutations. Cellular misfunctioning, cell death, aging, and cancer are the phenotypical consequences of DNA damages accumulation in the genome. Fortunately, an intricate set of genome surveillance mechanisms function to counteract genomic insults. Among these mechanisms, base excision repair and nucleotide excision repair are both dedicated to the removal of single-strand lesions contrary to double-strand break repair. Additionally, some specialized polymerases can temporarily take over lesion-arrested DNA polymerases during S-phase, in a mutagenic mechanism called translesion synthesis. Such polymerases only work if a more reliable system, such as homologous recombination, cannot avoid stumbled DNA replication. These DNA repair mechanisms function in conjunction with an intricate machinery of damage sensors, responsible of a series of phosphorylations and chromatin modifications that signal to the rest of the cell the presence of lesions on the DNA. Together DNA repair mechanisms and DNA damage signaling systems form a molecular shield against genomic instability called DNA Damage Response system. Hence, we have tried to integrate several papers that present a synergy that emerge when researchers from different fields put their forces together into a common goal, trying to improve human health. We thank the contributors for their work and the many reviewers who served conscientiously and tirelessly to assure an issue that meets the standards.

Frédéric Coin
Bernardo Reina-San-Martin
Giuseppina Giglia-Mari
Mark Berneburg

Hindawi Publishing Corporation
Molecular Biology International
Volume 2012, Article ID 598984, 10 pages
doi:10.1155/2012/598984

Review Article

Genotoxicity Studies Performed in the Ecuadorian Population

César Paz-y-Miño, Nadia Cumbal, and María Eugenia Sánchez

Instituto de Investigaciones Biomédicas, Facultad de Ciencias de la Salud, Universidad de las Américas, Ave. de los Granados y Colimes Quito, 1712842, Ecuador

Correspondence should be addressed to César Paz-y-Miño, cpazymino@udla.edu.ec

Received 3 February 2011; Revised 25 November 2011; Accepted 5 December 2011

Academic Editor: Mark Berneburg

Genotoxicity studies in Ecuador have been carried out during the past two decades. The focuses of the research were mainly the area of environmental issues, where the populations have been accidentally exposed to contaminants and the area of occupational exposure of individuals at the workplace. This paper includes studies carried out in the population of the Amazon region, a zone known for its rich biodiversity as well as for the ecological damage caused by oil spills and chemical sprayings whose consequences continue to be controversial. Additionally, we show the results of studies comprised of individuals occupationally exposed to toxic agents in two very different settings: flower plantation workers exposed to pesticide mixtures and X-ray exposure of hospital workers. The results from these studies confirm that genotoxicity studies can help evaluate current conditions and prevent further damage in the populations exposed to contaminants. As such, they are evidence of the need for biomonitoring employers at risk, stricter law enforcement regarding the use of pesticides, and increasingly conscientious oil extraction activities.

1. Introduction

Genotoxicity is a collective term that refers to any process that affects the structural integrity of DNA [1]. This multidisciplinary field of research aims to detect compounds capable of causing DNA damage in hopes of understanding the biological consequences of genotoxic agents and their involvement in the alteration of the molecular mechanisms of the genetic material [2]. These consequences can eventually lead to carcinogenic processes [3]. Over the past century, industrialization and globalization of the western hemisphere lead to the high volume production of different chemicals and complex preparations that are still currently released into the environment [4]. Living organisms are increasingly being exposed to genotoxic agents whose growing presence in the biosphere can substantially harm the population [5]. Activities such as fuel extraction and glyphosate spraying in the Amazon region of Ecuador are the two most controversial environmental health issues in the nation and are still considered as latent threats whose consequences continue to be studied [6, 7]. Agriculture in Ecuador is the second most important productive activity that contributes to national income [8]. However, the lack of regulation regarding pesticide use and occupational safety pose a significant threat to the workers' health [9]. Additionally, various studies have focused on individuals exposed to radiation in the workplace, such as medical radiation workers constantly exposed to ionizing radiation that has well-known DNA-damaging effects [10–12]. The present paper intends to show a summary on the work carried out in Ecuador for the past two decades in the field of genotoxicity. All the cytogenetic studies have been performed on blood lymphocytes and the results obtained refer only to somatic mutations. The studies have included cytogenetic findings, such as the ones presented in Table 1 as well as molecular results which are shown in Table 2.

2. Glyphosate Genotoxicity Studies

The northeastern Ecuadorian border underwent the aerial spraying of an herbicide mix during the period of 2002–2007 and was supported by the Colombian government [13]. The Roundup mix presumably contained high doses of glyphosate plus a surfactant known as polyethoxylated tallowamine (POEA) and the adjuvant Cosmoflux 411F [14].

Table 1: Cytogenetic findings in genotoxicity studies.

Glyphosate	Paz-y-Miño et al. [6]	Comet assay: 35.5 μm DNA migration for exposed, 25.94 μm for controls.
	Paz-y-Miño et al. [13]	All the studied population showed low or no chromosomal fragility.
Other pesticides	Paz-y-Miño et al. [9, 10, 93]	Chromosomal aberrations: 20.59% in exposed and 2.73% in controls.
	Paz-y-Miño et al. [55]	Comet assay: 31.58 μm DNA migration for exposed, 25.94 μm for controls. Chromosomal aberrations: 5.48% in exposed and 0.45% in controls.
Hydrocarbons	Paz-y-Miño et al. [7]	Chromosomal aberrations: 20% in exposed and 1-2% in controls. 12% type A DNA damage and 1% type E DNA damage in exposed group while 81% type A and 0% type E in controls.
	Paz-y-Miño et al. [88]	48.8% type A DNA damage and 0.1% type E DNA damage in exposed group while 67.9% type A and 0% type E in controls.
Radiation	Paz-y-Miño et al. [102]	Chromosomal aberrations: 29% in exposed, 26.0% in the followup, and 3.5% in controls.
	Paz-y-Miño et al. [112]	12.6% metaphases with telomeric associations in the exposed smoker group; 6.0% TA in exposed nonsmokers; 9.0% in unexposed smokers and 0.1% TA in control group.
	Paz-y-Miño et al. [9, 10, 93]	Comet assay: 26.55 μm DNA migration; mean chromosomal aberrations without gaps: 5.39% ($r = 0.50$, $P < 0.05$). Mean chromosomal aberrations including gaps: 12.08% ($r = 0.78$, $P < 0.01$).
	Muñoz et al. [128]	Comet assay: 29.08 μm DNA migration for exposed group, 25.91 μm for controls. Chromosomal aberrations: 50% in exposed, 26.0% in the followup, and 4% in controls.

Table 2: Molecular findings in genotoxicity studies.

Glyphosate	Paz-y-Miño et al. [13]	Regarding the *GSTP1* Ile105Val polymorphism, the frequency of the Val allele was higher in exposed individuals (0.48) than control individuals (0.28). The Val/Val variant represented a 4.88-fold risk of acquiring detoxification problems, whereas the combination of the Ile/Val and Val/Val alleles was associated with a 2.6-fold risk of presenting a *GSTP1* gene dysfunction. As for the *GPX-1* Pro198Leu polymorphism, the Leu allele had a higher frequency in exposed individuals (0.41), unlike control individuals (0.32). The Leu/Leu variant was associated with an 8.5-fold risk of having problems in the function of the *GPX-1* gene.
Other pesticides	Paz-y-Miño et al. [55]	The level of damage was not significantly influenced by genetic polymorphisms of the *CYP 1A1* gene in the studied population.
Hydrocarbons	Paz-y-Miño et al. [7]	As far as the *MSH2* gene is concerned, there is a relation between polymorphisms of the exon 13 and the DNA damage evaluated in the individuals exposed to hydrocarbons ($P < 0.001$), the study of the *CYP 1A1* gene found no relation between its polymorphisms and having greater susceptibility to DNA damage.

Glyphosate is an effective organophosphorous herbicide used worldwide [15, 16] known to cause variable levels of toxicity in different organisms, such as the alteration of metabolic pathways, cytotoxicity in humans, metamorphosis alterations in amphibians, and abnormal development of sea urchin eggs [17–22]. Such reports add to the already numerous concerns over the compound's many effects in the environment. The cytogenetic study of blood lymphocytes from individuals that lived in the area that endured the sprayings showed the absence of chromosomal aberrations two years after the last spraying in Ecuadorian soil [13]. Also, since glyphosate has been known to cause oxidative stress in microorganisms, plant and animal species [23–29], the study analyzed three gene polymorphisms (*GSTP1* Ile105Val, *GPX-1* Pro198Leu, and *XRCC1* Arg399Gln) that have been previously associated to the alteration of antioxidant activity, DNA detoxicating processes, and protective functions [30–33]. The *GSTP1* gene encodes for glutathione S-transferase pi, an enzyme that is involved in the protection against exogenous and endogenous oxidative damage [34]. As a member of the gluthatione-S-transferase superfamily of enzymes, GSTpi participates in the conjugation of xenobiotics, such as herbicides, insecticides, and other environmental carcinogens, to form glutathione and facilitate their excretion [35, 36]. Specifically, the *GSTP1* Ile105Val polymorphism has been associated to higher levels of DNA damage in pesticide-exposed populations [37]. In the Ecuadorian population studied, the prevalence of the *GSTP1* Val/Val genotype associated with enzyme dysfunction was observed in the exposed individuals. On the other hand, the *GPX-1* gene encodes for one of the most important detoxifying enzymes: gluthatione peroxidase. This enzyme protects mammalian cells, especially human erythrocytes, against oxidative damage [38]. Studies have shown that the loss of gluthatione peroxidase activity can generate tissue damage [39] and offer more sensitivity towards toxic

xenobiotics, such as paraquat and adriamycin [38]. Although the *GPX-1* Pro198Leu polymorphism has been exhaustively studied in relation to cancer, the study in Ecuador identified the prevalence of the Leu allele of the *GPX-1* gene in glyphosate-exposed individuals that suggests a higher risk of DNA damage and increased sensitivity to herbicides. The third gene that was part of the study, *XRCC1*, is involved in the mechanisms of DNA single-strand breaks (SSBs) and base-excision repair (BER) that could modify the individual susceptibility to the genotoxic effect of xenobiotics [40, 41]. Although other studied populations have found an association between *XRCC1* gene genotypes and an increased risk of DNA damage due to pesticide exposure, similar results were not found for the Ecuadorian population studied [42]. Aside from the cytogenetic and molecular analysis carried out, the social conditions of the population were surveyed and psychological assessment was offered to all the individuals. The results of these two activities suggested the negative effect of the fumigations on the individuals' mental health, social conditions, and quality of life [13].

A previous study took place two years before the aforementioned study at the Ecuadorian border. It involved individuals living within 200 m to 3 Km from the areas under continuous and sporadic spraying [14, 43]. The comet assay technique, described by Singh et al., 1988 [44], was carried out on blood samples from exposed individuals and corresponding controls to show the occurrence of DNA fragmentation. DNA damage was classified into five categories and the mean of DNA migration was recorded. The results showed that the exposed group displayed significantly higher mean DNA migration than the control group. Similarly, there was a higher degree of DNA damage in the exposed group in comparison to the control group. These results suggest a negative effect of the glyphosate formulation since none of the studied individuals had been previously or simultaneously exposed to other toxic compounds, such as pesticides or tobacco [6]. The northern strip on the Ecuadorian border has not gone unnoticed in the controversy regarding aerial sprayings and their consequences. Nevertheless, the genotoxic and overall toxic potential of glyphosate remains under study and in vitro findings [45–49] have reached a variety of results [29]. The two studies carried out in the Ecuadorian border suggest a significant and immediate risk arising from the use of this chemical and prompt to continue its investigation.

3. Other Pesticide Genotoxicity Studies

Despite the known risks of the use of some pesticides due to their potential health consequences [25], many of those catalogued as extremely toxic continue being used in certain agricultural zones and flower plantations in Ecuador [9]. Pesticides are widely used all over the world in agriculture to protect crops and in public health to control diseases [50, 51]. The risk of developing malignancies such as cancer in occupationally exposed populations is of great concern and has drawn attention to workers in various activities, from the manufacturing workers to the pesticide applicators [52]. Studies available in scientific literature have focused their methodology on cytogenetic endpoints to evaluate the potential genotoxicity of pesticides, including chromosomal aberrations (CAs), micronuclei (MN), and sister chromatid exchanges (SCEs) [52, 53]. The analysis of chromosomal aberrations such as breaks, dicentric chromosomes, and rings was part of the methodology used in a leading pesticide exposure study carried out in flower plantation workers in Quito. These workers were exposed to 27 different pesticides, some of which have been previously labeled as highly toxic [9]. Also, the level of erythrocyte acetylcholinesterase was measures in every individual as a marker to evaluate the exposure to organophosphate pesticides [54]. The study found an overall CA percentage of 20.59% in the exposed group and 2.73% of CA in the control group. Additionally, the exposed group showed a higher proportion of chromatid-type aberrations and numeric alterations. This does not only reflect genomic instability but also comprises outstanding evidence of damage supported by the abnormal low levels of acetylcholinesterase seen in the exposed group [55].

Pesticide genotoxicity was also studied in individuals working as pesticide applicators in the zone of Cayambe, northeast from Quito. The workers were exposed to 46 pesticides of different degrees of toxicity and at different concentrations and mixtures during work at the plantation [56]. The methodology involved chromosomal aberration test matched with alkaline comet assay [55]. Additionally, the samples were analyzed at a molecular level focusing on the *CYP1A1* gene, a gene that has been extensively studied in relation to occupationally exposure to pesticides [56]. Because the gene is involved in the human xenobiotic metabolism, its alteration presumably increases the risk of developing lung, colorectal, prostate, and breast cancer [57–61]. In accordance to the first study carried out in workers from Quito, the results of the CA analysis in this study showed the significantly high presence of chromosomal damage in the exposed group as compared to the control group. Furthermore, the comet assay test offered results that supported the CA analysis by showing that the DNA migration of the exposed group was certainly higher than that of the control group. On the other hand, the study presented no correlation between the cytogenetic findings and genotyping of the *CYP1A1* MspI and Ile/Val gene polymorphisms. Though CA and comet assay showed interesting results, the gene was not linked to pesticide exposure in the studied population, as opposed to other populations [62, 63].

As an important element of the agricultural production, pesticides have become a necessary tool for crop management in developing countries [64, 65]. Nonetheless, the lack of adequate legislation and enforcement of existing pesticide laws and regulations places agricultural workers, their families, and nearby populations in great risk of developing cancer and other diseases [66]. Our study shows evidence of genotoxic damage in individuals occupationally exposed to pesticides in Ecuador. These are results that demand the establishment of effective exposure biomarkers that could be used for biomonitoring the threatened workers in order to prevent the future development of illness [5].

4. Hydrocarbons Genotoxicity Studies

The Ecuadorian Amazon is one of the ecologically richest regions in the world and it is also sparsely populated. The oil extraction activity in Ecuador began in 1972, it became economically fundamental immediately and continues to be the principal source of national income [67]. Unfortunately, along the process, millions of gallons of oil and toxic residues have been discarded directly onto the environment causing health and environmental issues [68–71]. Indeed, more than 30 billion gallons of toxic wastes and crude oil had been discharged into the land and waterways of the Ecuadorian Amazon up until 1993 [72]. Crude oil is a complex mixture of many chemical compounds. It contains a variety of hydrocarbons of diverse toxicological power such as benzene, toluene, xylene and polynuclear aromatic hydrocarbons [73]. High concentrations of benzene can cause neurotoxin symptoms that cause injuries to the bone marrow and, less frequently, pancytopenia [74]. Similarly, benzene is known to cause leukemia and the development of hematological tumors [75]. The exposure to carcinogen compounds used in the oil industry increases the development of cancer in men, women, and children. In men, an increase of lung, esophagus, rectum, skin, and kidney cancer has been noticed. In women, researchers have seen an increase of cervical, lymphatic ganglion, and bladder cancer. In children, an increase of hematopoietic cancer has been shown among other types of cancer [75–86]. Studies carried out in the Ecuadorian Amazon Basin were found to be compatible with international studies. A relationship between cancer incidence and living in proximity to oil fields has been established [87]. An initial study was carried out in the province of Orellana including 23 women living not more than 10 Km away from a crude oil extraction zone, with the corresponding control group. In order to assess genotoxicity, the comet assay test was used to measure DNA damage by classifying the nucleus morphology into five categories. The sampling zones closest to the extraction wells showed a greater evidence of DNA damage than those that are farther away which suggests a distance-damage relationship. This relationship is also supported by the increased occurrence of type A nuclei (no damage) as the distance from the wells increases [88]. Another study was carried out with a significantly bigger sample size from the nearby zone of San Carlos and matching controls from both San Carlos and the country's capital Quito. Comet assay showed that the affected group has a high occurrence of type B nuclei fragmentation, as opposed to the prevalence of type A cells in the control groups from San Carlos and Quito. Additionally, the analysis of chromosomal aberrations showed that 20% of the exposed individuals presented chromosomal breaks and gaps while only 2% of the control individuals had such aberrations [88]. At a molecular level, we incorporated the analysis of the polymorphisms of the genes *CYP1A1* (MspI and Ile/Val) and *MSH2* (gIVS12-6T>C), both related to the development of cancer [58, 59, 89, 90]. The results of this last part of the study showed that the *CYP1A1* gene polymorphisms were not related to either group, as it has been reported previously in the Caucasian population [58]. However, the study showed a significant difference of the *MSH2* gene polymorphism between groups which suggests a higher susceptibility to DNA damage in the exposed group [7]. The chemical complexity of petroleum causes that, once a spill occurs, the constituents disseminate into different extents between the oil phase and the air, soil and water phases of the environment. Physical, chemical, and biological processes age the spilled product resulting in additional changes in composition and complexity [91]. Taking in to consideration the toxicity of these fractions, the risk at petroleum extraction sites is an issue that must be addressed by making informed decisions. By comparing the affected group with the control individuals living in the same town though far from the extraction sites, the study has been able to evidence of the genotoxicity in the exposed population living in nearby petroleum extraction wells. This suggests that the contaminating material resulting from this activity has created an altered environment that exposes the population to chemical fractions considered as dangerous and may also cause genotoxic effects.

5. Radiation Genotoxicity Studies

Environmental mutagens can be broadly classified as radiation and chemicals [92]. Ionizing radiation is capable of extracting electrons of the radiated material due to its high energy. This is only a start point for other ionizing reactions that produce more unstable molecules that eventually cause mutations in DNA [93].

Ever since X-rays were shown to induce mutation in *Drosophila* over 70 years ago, the established idea has been that the genotoxic effects of ionizing radiation, such as mutations and carcinogenesis, are caused by the direct damage of the cell nucleus [94, 95]. Diagnostic radiology is a field of physical medicine that uses X-rays in order to obtain functional and anatomical information on the human body [96]. Because of the benefits of this diagnostic tool that allows real-time visualization, it is frequently used by the medical professionals [96].

Ionizing radiation is capable of acting on the living cell causing several effects that result from the excitation of atoms and molecules that ultimately cause structural changes. At a molecular level, DNA is possibly affected by water ionization that forms free radicals and promotes the oxidation of several compounds and hydrogen oxide [97]. Even small doses of this radiation could cause great damage because a simple electron excitation can break up to 20 hydrogen bonds [98–100]. The damage resulting from radiation exposure can be seen in the form of chromosomal aberrations in the cell nucleus that are associated with an elevating risk of developing cancer [101]. A first study focused on 10 individuals exposed to radiation in the workplace. They were exposed to 1.84 mSv/year and received a follow-up cytogenetic study after a year from the first blood sampling. The chromosomal aberration results found by the cytogenetic analysis showed interesting results in both instances [102]. First, they showed that complex chromosome alterations, such as rings and dicentrics, are

present in low percentages contrary to the occurrence of simple alterations (gaps, breaks, and acentrics). This is due to the fact that 72 h cultures were used. By then, cells have gone through a second and third mitotic divisions. Therefore, primary alterations have been kept and turned into secondary aberrations in the growing generations; meanwhile, early cells with complex alterations have already died [103, 104]. Two individuals exposed to higher doses of radiation (4.54 and 1.07 mSv, resp.) did show complex alterations in the second sampling. This is an unusual finding likely to be caused by the individuals' sporadic exposure to higher amounts of radiation. A significant increase of chromosomal aberrations was observed by comparing CA during the first and the second sampling. Out of these, there was a higher number of lesions at the chromatid level (mostly gaps) possibly due to the proper action of DNA repair mechanisms at low doses of exposure over long periods of time [102, 105]. Other studies have also found an increase of CA in individuals exposed to similar doses of radiation, but have not addressed the importance of periodic biomonitoring [106]. Though numerical aberrations were not the focus of the study, the exposed individuals showed an increased frequency of hyperploidies and hypoploidies possibly due to the imbalance in the cell cycle caused by the exposure to toxic agents [107]. This data evidences the importance of periodic control of the occupational exposure and of monitoring the dosage-time levels of exposure at the workplace [102].

A second study focused on telomeric associations in individuals exposed both to X-rays and smoke in order to determine the existence of these associations as chromosomal markers of exposure to these carcinogenic agents. The phenomenon of telomeric association is an intermediate step in the progression towards chromosomal instability that also comprises a risk of developing cancer [108, 109]. Cytogenetic monitoring is currently accepted as an evaluation tool for exposed populations at risk as it has been used in studies regarding ionizing radiation [5]. Cytogenetic analysis of cigarette smokers has shown the occurrence of chromosomal aberrations in populations from Colombia and India [110, 111]. In this study, mitotic indices determined in all groups (smokers exposed to radiation, nonsmokers exposed to radiation, smokers unexposed, and unexposed nonsmokers) showed no correlations between the exposure to both carcinogens and the mitotic indices and cell proliferation. Nonetheless, the three different exposed groups showed high frequencies of telomeric associations [112]. These results were particularly surprising for cigarette smokers since no cytogenetic biomarker of exposure for cigarette had been demonstrated to be consistent so far. Also, the group of smokers unexposed to radiation showed higher frequencies of TA than the nonsmoking X-ray-exposed group, a result that was possibly due to the many carcinogens present in cigarette smoke [110, 111, 113]. Though it has been reported that both agents have a synergistic effect [113], our study did not find such a tendency. However, the group exposed to both agents did show the highest frequency of telomeric associations. The study suggested that telomeric associations can assess the genomic instability phenomena in populations exposed to mutagens. Although, telomere length has been reported as a biomarker for age, stress and cancer, telomere biology and the molecular pathways that protect telomeres continue to be studied in order to determining the outcome of radiation exposure [114].

In another study, the inclusion of gaps as chromosomal aberration was investigated. Gaps are defined as the unstained regions of a chromosome that contain zones of lesser width than that of a chromatid [115, 116]. A gap is observed as an empty space because the DNA thread is so thin that it becomes practically invisible to the usual technique [113]. Since genotoxic agents such as ionizing radiation are capable of inducing chromosomal uncoiling events and affecting DNA condensation, gaps can certainly be the product of exposure to genotoxic agents [117]. Nonetheless, some authors had considered gaps to be structures that lack biological significance [118]. The study involved individuals exposed to X-rays and the unexposed control group. The findings of the CA analysis and comet assay were compared, including and excluding gaps. These two complementary techniques do not detect the same kind of lesions. On one hand, chromosomal aberrations are originated from double-strand breaks; on the other hand, comet assay can detect single-strand breaks, double-strand breaks, and alkali labile sites (when using the alkaline version) [119] and has proven to be a useful way of assessing X-ray damage to lymphocytes [120]. The correlation between the two methods including gaps as CA was positive. Gaps measured damage in the DNA since there was a stronger correlation between the results of both applied techniques when gaps were included as a CA. Although there is an increasing interest in studying the more complex chromosomal aberrations such as dicentrics and translocations, current studies still include gaps as part of the genotoxicity studies [121–124]. These findings suggested a revision of the biological importance of gaps in population occupational biomonitoring [124]. Furthermore, another study involved a group of radiologists and technicians exposed to X-rays at the workplace, excluding those with family and personal history of cancer and smoke exposure. The mean dose of ionizing radiation for the affected group was 0.99 mSv and the chromosomal aberrations observed involved gaps, breaks, dicentric, rings, and double minutes. The cytogenetic analysis showed that CAs were present in 50% of the individuals in the exposed group and in 2% of the control individuals. However, these results were not statistically significant. On the other hand, the comet assay did show a highly significant difference of migration in the exposed group as compared to the control group, possibly because the comet assay shows a wider set of damage consequences [93, 125]. Similarl to a study carried out in Iran that found no specific relation with the characteristics of the occupational setting and the duration of exposure [126], this study found no correlation between the results of both tests and the duration and dose of exposure due to a lack of significant variation between individual doses. Such results may also support the idea of hormesis taking place as a way of adapting to the workplace after several years, though the idea remains to be controversial [127]. Even though there were no significant results regarding CA, relative risk calculation

showed that exposed individuals had a risk 20 times higher of showing aberrations than the control group [128]. These aberrations may lead to the alteration of cell control mechanisms such as apoptosis and tumor suppressor genes, besides the loss of genetic material due to cell death as a result of changes in division and repairing mechanisms [129–131]. Although this study did not show an association between CA and exposure time or dose level, other studies have agreed on the fact that long-term exposure to low radiation levels are the cause for higher percentages of CA [132, 133].

Cytogenetic findings are of great importance because they are associated with the mechanisms of carcinogenesis. The interaction with physical agents, such as ionizing radiation, produces a variety of primary lesions [134] whose prevalence can determine cancer risk [135]. Because of the importance of biomonitoring occupationally exposed populations, proper research guidelines have been established [136]. mFISH assays are currently being put in use in order to carry out a more detailed analysis of simple and complex aberrations that could model the effects of radiation on lymphocytes [107]. Nonetheless, earlier cytogenetic techniques are still held as the golden standard for biomonitoring populations. Environmental and occupational health issues are increasingly being studied because of its importance in public health. In Ecuador, going through a preliminary cytogenetic testing to evaluate the genotoxic effects of different agents is a rather voluntary decision and the toxic qualities of certain widely used chemicals, such as herbicides and pesticides, are not of common knowledge. According to the results obtained in this set of studies, adequate biomonitoring laws should be enforced. In the case of glyphosate, research has helped to consider changes regarding the targeted areas, duration of sprayings and chemical composition of herbicides. Though plantations and industries do offer protective equipment, this gear is not always used by the small farm owner. Reports on the use of highly toxic pesticides conclude that the lack of regulation of pesticide use benefits the informal distribution of these hazardous compounds, not only in flower plantations but increasingly in small farms. Lastly, occupational health risks must be studied in all professions that face any level of exposure to physical or chemical agents suspected to cause illness. Radiologists and other professionals exposed to radiation should have access to cytogenetics testing and follow-up studies that can report on any unusual results in order to prevent diseases as part of occupational health and safety laws.

References

[1] J. Bohne and T. Cathomen, "Genotoxicity in gene therapy: an account of vector integration and designer nucleases," *Current Opinion in Molecular Therapeutics*, vol. 10, no. 3, pp. 214–223, 2008.

[2] M. Uhl, M. J. Plewa, B. J. Majer, and S. Knasmüller, "Basic principles of genetic toxicology with an emphasis on plant bioassays," in *Bioassays in Plant Cells for Improvement of Ecosystem and Human Health*, J. Maluszynska and M. Plewa, Eds., pp. 11–30, Katowice, Poland, 2003.

[3] G. H. Westphalen, L. M. Menezes, D. Prá et al., "In vivo determination of genotoxicity induced by metals from orthodontic appliances using micronucleus and comet assays," *Genetics and Molecular Research*, vol. 7, no. 4, pp. 1259–1266, 2008.

[4] F. Henkler and A. Luch, "Adverse health effects of environmental chemical agents through non-genotoxic mechanisms," *Journal of Epidemiology and Community Health*, vol. 65, no. 1, pp. 1–3, 2011.

[5] T. Adamus, I. Mikulenková, L. Dobiáš, J. Havránková, and T. Pek, "Cytogenetic methods and biomonitoring of occupational exposure to genotoxic factors," *Journal of Applied Biomedicine*, vol. 4, no. 4, pp. 197–203, 2006.

[6] C. Paz-y-Miño, M. E. Sánchez, M. Arévalo et al., "Evaluation of DNA damage in an Ecuadorian population exposed to glyphosate," *Genetics and Molecular Biology*, vol. 30, no. 2, pp. 456–460, 2007.

[7] C. Paz-y-Miño, A. López-Cortés, M. Arévalo, and M. E. Sánchez, "Monitoring of DNA damage in individuals exposed to petroleum hydrocarbons in Ecuador," *Annals of the New York Academy of Sciences*, vol. 1140, pp. 121–128, 2008.

[8] A. Joaquín, S. Vallejo, and R. Trejos, *Más que Alimentos en la Mesa: La Real Contribución de la Agricultura a la Economía del Ecuador*, vol. 11, Instituto Interamericano de Cooperación para la Agricultura (IICA), Quito, Ecuador, 2005.

[9] C. Paz-y-Miño, G. Bustamante, M. E. Sáchez, and P. E. Leone, "Cytogenetic monitoring in a population occupationally exposed to pesticides in ecuador," *Environmental Health Perspectives*, vol. 110, no. 11, pp. 1077–1080, 2002.

[10] C. Paz-y-Miño, M. V. Dávalos, M. E. Sánchez, M. Arévalo, and P. E. Leone, "Should gaps be included in chromosomal aberration analysis?: evidence based on the comet assay," *Mutation Research*, vol. 516, no. 1-2, pp. 57–61, 2002.

[11] M. Stoia, S. Oancea, and D. C. Obreja, "Comparative study of genotoxic effects in workers exposed to inorganic lead and low dose irradiation using micronucleus test," *Romanian Journal of Legal Medicine*, vol. 17, no. 4, pp. 287–294, 2009.

[12] V. Garaj-Vrhovac and N. Kopjar, "The alkaline Comet assay as biomarker in assessment of DNA damage in medical personnel occupationally exposed to ionizing radiation," *Mutagenesis*, vol. 18, no. 3, pp. 265–271, 2003.

[13] C. Paz-y-Miño, M. J. Muñoz, A. Maldonado et al., "Baseline determination in social, health, and genetic areas in communities affected by glyphosate aerial spraying on the northeastern Ecuadorian border," *Reviews on Environmental Health*, vol. 26, no. 1, pp. 45–51, 2011.

[14] Ministerio de Relaciones Exteriores (MREE), *Misión de Verificación: Impactos en el Ecuador de las Fumigaciones Realizadas en el Departamento del Putumayo dentro del Plan Colombia*, Ministerio de Relaciones Exteriores del Ecuador, Quito, Ecuador, 2002.

[15] S. O. Duke and S. B. Powles, "Glyphosate: a once-in-a-century herbicide," *Pest Management Science*, vol. 64, no. 4, pp. 319–325, 2008.

[16] J. F. Acquavella, B. H. Alexander, J. S. Mandel, C. Gustin, B. Baker, and P. Chapman, "Glyphosate biomonitoring for farmers and their families: results from the farm family exposure study," *Environmental Health Perspectives*, vol. 112, no. 3, pp. 321–326, 2004.

[17] N. Benachour and G. E. Séralini, "Glyphosate formulations induce apoptosis and necrosis in human umbilical, embryonic, and placental cells," *Chemical Research in Toxicology*, vol. 22, no. 1, pp. 97–105, 2009.

[18] A. Martínez, I. Reyes, and N. Reyes, "Cytotoxicity of the herbicide glyphosate in human peripheral blood mononuclear cells," *Biomedica*, vol. 27, no. 4, pp. 594–604, 2007.

[19] C. M. Howe, M. Berrill, B. D. Pauli, C. C. Helbing, K. Werry, and N. Veldhoen, "Toxicity of glyphosate-based pesticides to four North American frog species," *Environmental Toxicology and Chemistry*, vol. 23, no. 8, pp. 1928–1938, 2004.

[20] S. K. Dinehart, L. M. Smith, S. T. McMurry, T. A. Anderson, P. N. Smith, and D. A. Haukos, "Toxicity of a glufosinate- and several glyphosate-based herbicides to juvenile amphibians from the Southern High Plains, USA," *Science of the Total Environment*, vol. 407, no. 3, pp. 1065–1071, 2009.

[21] J. Marc, O. Mulner-Lorillon, S. Boulben, D. Hureau, G. Durand, and R. Bellé, "Pesticide roundup provokes cell division dysfunction at the level of CDK1/cyclin B activation," *Chemical Research in Toxicology*, vol. 15, no. 3, pp. 326–331, 2002.

[22] R. Bellé, R. Le Bouffant, J. Morales, B. Cosson, P. Cormier, and O. Mulner-Lorillon, "Sea urchin embryo, DNA-damaged cell cycle checkpoint and the mechanisms initiating cancer development," *Journal de la Societe de Biologie*, vol. 201, no. 3, pp. 317–327, 2007.

[23] D. M. Romero, M. C. Ríos de Molina, and Á. B. Juárez, "Oxidative stress induced by a commercial glyphosate formulation in a tolerant strain of Chlorella kessleri," *Ecotoxicology and Environmental Safety*, vol. 74, no. 4, pp. 741–747, 2011.

[24] N. S. El-Shenawy, "Oxidative stress responses of rats exposed to Roundup and its active ingredient glyphosate," *Environmental Toxicology and Pharmacology*, vol. 28, no. 3, pp. 379–385, 2009.

[25] M. Mladinic, S. Berend, A. L. Vrdoljak, N. Kopjar, B. Radic, and D. Zeljezic, "Evaluation of genome damage and its relation to oxidative stress induced by glyphosate in human lymphocytes in vitro," *Environmental and Molecular Mutagenesis*, vol. 50, no. 9, pp. 800–807, 2009.

[26] O. V. Lushchak, O. I. Kubrak, J. M. Storey, K. B. Storey, and V. I. Lushchak, "Low toxic herbicide Roundup induces mild oxidative stress in goldfish tissues," *Chemosphere*, vol. 76, no. 7, pp. 932–937, 2009.

[27] K. A. Modesto and C. B. R. Martinez, "Roundup® causes oxidative stress in liver and inhibits acetylcholinesterase in muscle and brain of the fish Prochilodus lineatus," *Chemosphere*, vol. 78, no. 3, pp. 294–299, 2010.

[28] N. Ahsan, D. G. Lee, K. W. Lee et al., "Glyphosate-induced oxidative stress in rice leaves revealed by proteomic approach," *Plant Physiology and Biochemistry*, vol. 46, no. 12, pp. 1062–1070, 2008.

[29] L. Goldman, *Childhood Pesticide Poisoning*, United Nations Environment Program, Geneva, Switzerland, 2004, http://www.who.int/ceh/publications/pestpoisoning.pdf.

[30] I. Meiers, J. H. Shanks, and D. G. Bostwick, "Glutathione S-transferase pi (GSTP1) hypermethylation in prostate cancer: review 2007," *Pathology*, vol. 39, no. 3, pp. 299–304, 2007.

[31] H. W. Lo, L. Stephenson, X. Cao, M. Milas, R. Pollock, and F. Ali-Osman, "Identification and functional characterization of the human Glutathione S-transferase P1 gene as a novel transcriptional target of the p53 tumor suppressor gene," *Molecular Cancer Research*, vol. 6, no. 5, pp. 843–850, 2008.

[32] A. M. Moyer, O. E. Salavaggione, T. Y. Wu et al., "Glutathione S-transferase P1: gene sequence variation and functional genomic studies," *Cancer Research*, vol. 68, no. 12, pp. 4791–4801, 2008.

[33] R. H. Wong, C. L. Du, J. D. Wang, C. C. Chan, J. C. J. Luo, and T. J. Cheng, "XRCC1 and CYP2E1 polymorphisms as susceptibility factors of plasma mutant p53 protein and anti-p53 antibody expression in vinyl chloride monomer-exposed polyvinyl chloride workers," *Cancer Epidemiology Biomarkers and Prevention*, vol. 11, no. 5, pp. 475–482, 2002.

[34] C. Jerónimo, G. Varzim, R. Henrique et al., "I105V polymorphism and promoter methylation of the GSTP1 gene in prostate adenocarcinoma," *Cancer Epidemiology Biomarkers and Prevention*, vol. 11, no. 5, pp. 445–450, 2002.

[35] L. Kadouri, Z. Kote-Jarai, A. Hubert et al., "Glutathione-S-transferase M1, T1 and P1 polymorphisms, and breast cancer risk, in BRCA1/2 mutation carriers," *British Journal of Cancer*, vol. 98, no. 12, pp. 2006–2010, 2008.

[36] C. Martínez, E. García-Martín, H. Alonso-Navarro et al., "Glutathione-S-transferase P1 polymorphism and risk for essential tremor," *European Journal of Neurology*, vol. 15, no. 3, pp. 234–238, 2008.

[37] Y. J. Liu, P. L. Huang, Y. F. Chang et al., "GSTP1 genetic polymorphism is associated with a higher risk of DNA damage in pesticide-exposed fruit growers," *Cancer Epidemiology Biomarkers and Prevention*, vol. 15, no. 4, pp. 659–666, 2006.

[38] J. M. Matés, "Effects of antioxidant enzymes in the molecular control of reactive oxygen species toxicology," *Toxicology*, vol. 153, no. 1–3, pp. 83–104, 2000.

[39] J. R. Arthur, "The glutathione peroxidases," *Cellular and Molecular Life Sciences*, vol. 57, no. 13-14, pp. 1825–1835, 2000.

[40] P. Rohr, J. da Silva, B. Erdtmann et al., "BER gene polymorphisms (OGG1 Ser326Cys and XRCC1 Arg194Trp) and modulation of DNA damage due to pesticides exposure," *Environmental and Molecular Mutagenesis*, vol. 52, no. 1, pp. 20–27, 2011.

[41] M. C. Stern, D. M. Umbach, C. H. Van Gils, R. M. Lunn, and J. A. Taylor, "DNA repair gene XRCC1 polymorphisms, smoking, and bladder cancer risk," *Cancer Epidemiology Biomarkers and Prevention*, vol. 10, no. 2, pp. 125–131, 2001.

[42] R. H. Wong, S. Y. Chang, S. W. Ho et al., "Polymorphisms in metabolic GSTP1 and DNA-repair XRCC1 genes with an increased risk of DNA damage in pesticide-exposed fruit growers," *Mutation Research*, vol. 654, no. 2, pp. 168–175, 2008.

[43] Acción Ecológica, *Frontera: Daños Genéticos Por las Fumigaciones del Plan Colombia*, Acción Ecológica, Quito, Ecuador, 2004.

[44] N. P. Singh, M. T. McCoy, R. R. Tice, and E. L. Schneider, "A simple technique for quantitation of low levels of DNA damage in individual cells," *Experimental Cell Research*, vol. 175, no. 1, pp. 184–191, 1988.

[45] D. D. Evans and M. J. Batty, "Effects of high dietary concentrations of glyphosate (Roundup®) on a species of bird, marsupial and rodent indigenous to Australia," *Environmental Toxicology and Chemistry*, vol. 5, no. 4, pp. 399–401, 1986.

[46] G. M. Williams, R. Kroes, and I. C. Munro, "Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans," *Regulatory Toxicology and Pharmacology*, vol. 31, no. 2, pp. 117–165, 2000.

[47] D. A. Goldstein, J. F. Acquavella, R. M. Mannion, and D. R. Farmer, "An analysis of glyphosate data from the California Environmental Protection Agency pesticide illness surveillance program," *Journal of Toxicology*, vol. 40, no. 7, pp. 885–892, 2002.

[48] M. H. Bernal, K. R. Solomon, and G. Carrasquilla, "Toxicity of formulated glyphosate (Glyphos) and cosmo-flux to larval and juvenile colombian frogs 2. field and laboratory microcosm acute toxicity," *Journal of Toxicology and Environmental Health—Part A*, vol. 72, no. 15-16, pp. 966–973, 2009.

[49] K. R. Solomon, A. Anadón, G. Carrasquilla, A. L. Cerdeira, J. Marshall, and L.-H. Sanin, "Coca and poppy eradication in Colombia: environmental and human health assessment of aerially applied glyphosate," *Reviews of Environmental Contamination and Toxicology*, vol. 190, pp. 43–125, 2007.

[50] N. Sailaja, M. Chandrasekhar, P. V. Rekhadevi et al., "Genotoxic evaluation of workers employed in pesticide production," *Mutation Research*, vol. 609, no. 1, pp. 74–80, 2006.

[51] S. Bull, K. Fletcher, A. R. Boobis, and J. M. Battershill, "Evidence for genotoxicity of pesticides in pesticide applicators: a review," *Mutagenesis*, vol. 21, no. 2, pp. 93–103, 2006.

[52] C. Bolognesi, "Genotoxicity of pesticides: a review of human biomonitoring studies," *Mutation Research*, vol. 543, no. 3, pp. 251–272, 2003.

[53] P. Grover, K. Danadevi, M. Mahboob, R. Rozati, B. S. Banu, and M. F. Rahman, "Evaluation of genetic damage in workers employed in pesticide production utilizing the Comet assay," *Mutagenesis*, vol. 18, no. 2, pp. 201–205, 2003.

[54] V. Ng, D. Koh, A. Wee, and S. E. Chia, "Salivary acetylcholinesterase as a biomarker for organophosphate exposure," *Occupational Medicine*, vol. 59, no. 2, pp. 120–122, 2009.

[55] C. Paz-y-Miño, M. Arévalo, M. E. Sanchez, and P. E. Leone, "Chromosome and DNA damage analysis in individuals occupationally exposed to pesticides with relation to genetic polymorphism for CYP 1A1gene in Ecuador," *Mutation Research*, vol. 562, no. 1-2, pp. 77–89, 2004.

[56] V. Kumar, C. S. Yadav, S. Singh et al., "CYP 1A1 polymorphism and organochlorine pesticides levels in the etiology of prostate cancer," *Chemosphere*, vol. 81, no. 4, pp. 464–468, 2010.

[57] A. M. Tsatsakis, A. Zafiropoulos, M. N. Tzatzarakis, G. N. Tzanakakis, and A. Kafatos, "Relation of PON1 and CYP1A1 genetic polymorphisms to clinical findings in a cross-sectional study of a Greek rural population professionally exposed to pesticides," *Toxicology Letters*, vol. 186, no. 1, pp. 66–72, 2009.

[58] C. San Jose, A. Cabanillas, J. Benitez, J. A. Carrillo, M. Jimenez, and G. Gervasini, "CYP1A1 gene polymorphisms increase lung cancer risk in a high-incidence region of Spain: a case control study," *BMC Cancer*, vol. 10, article 463, 2010.

[59] E. Taioli, L. Gaspari, S. Benhamou et al., "Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years," *International Journal of Epidemiology*, vol. 32, no. 1, pp. 60–63, 2003.

[60] J. Little, L. Sharp, L. F. Masson et al., "Colorectal cancer and genetic polymorphisms of CYP1A1, GSTM1 and GSTT1: a case-control study in the Grampian region of Scotland," *International Journal of Cancer*, vol. 119, no. 9, pp. 2155–2164, 2006.

[61] T. N. Sergentanis and K. P. Economopoulos, "Four polymorphisms in cytochrome P450 1A1 (CYP1A1) gene and breast cancer risk: a meta-analysis," *Breast Cancer Research and Treatment*, vol. 122, no. 2, pp. 459–469, 2010.

[62] V. M. Basham, P. D. P. Pharoah, C. S. Healey et al., "Polymorphisms in CYP1A1 and smoking: no association with breast cancer risk," *Carcinogenesis*, vol. 22, no. 11, pp. 1797–1800, 2001.

[63] K. Kvitko, J. C. B. Nunes, T. A. Weimer, F. M. Salzano, and M. H. Hutz, "Cytochrome P4501A1 polymorphisms in South American Indians," *Human Biology*, vol. 72, no. 6, pp. 1039–1043, 2000.

[64] A. V. Ngowi, *Health Impact of Exposure to Pesticides in Agriculture in Tanzania*, University of Tampere, Tampere, Finland, 2002.

[65] R. Naravaneni and K. Jamil, "Determination of AChE levels and genotoxic effects in farmers occupationally exposed to pesticides," *Human and Experimental Toxicology*, vol. 26, no. 9, pp. 723–731, 2007.

[66] IARC, *Evaluation of Carcinogenic Risks to Humans: Occupational Exposures in Insecticide Application and Some Pesticides*, vol. 53, The International Agency for Research on Cancer, Lyon, France, 1991, http://monographs.iarc.fr/ENG/Monographs/vol53/mono53.pdf.

[67] Banco Central del Ecuador (BCE), *Cifras Económicas del Ecuador Abril 2009*, Banco Central del Ecuador, Quito, Ecuador, 2009, http://www.bce.fin.ec/documentos/Estadisticas/SectorReal/Previsiones/IndCoyuntura/CifrasEconomicas/cie200904.pdf.

[68] P. R. Epstein and J. Selber, *A Life Cycle Analysis of Its Health and Environmental Impacts*, The Center for Health and the Global Environment, Boston, Mass, USA, 2002, http://chge.med.harvard.edu/publications/documents/oilfullreport.pdf.

[69] J. Kimmerling, *Amazon Crude*, Brickfron Graphics, New York, NY, USA, 1993.

[70] M. Neri, D. Ugolini, S. Bonassi et al., "Children's exposure to environmental pollutants and biomarkers of genetic damage: II. Results of a comprehensive literature search and meta-analysis," *Mutation Research*, vol. 612, no. 1, pp. 14–39, 2006.

[71] K. L. Platt, S. Aderhold, K. Kulpe, and M. Fickler, "Unexpected DNA damage caused by polycyclic aromatic hydrocarbons under standard laboratory conditions," *Mutation Research*, vol. 650, no. 2, pp. 96–103, 2008.

[72] C. Jochnick, R. Normand, and S. Zaidi, "Rights violations in the Ecuadorian Amazon: the human consequences of oil development," *Health & Human Rights*, vol. 1, no. 1, pp. 82–100, 1994.

[73] IARC, *Evaluation of the Carcinogenic Risk of Chemicals to Man: Occupational Exposures to Petroleum Refining; Crude Oil and Major Petroleum Fuels*, vol. 45, The International Agency for Research on Cancer, Lyon , France, 1989, http://monographs.iarc.fr/ENG/Monographs/vol45/mono45.pdf.

[74] I. Rahman, K. Narasimhan, S. Aziz, and W. Owens, "Gasoline ingestion: a rare cause of pancytopenia," *American Journal of the Medical Sciences*, vol. 338, no. 5, pp. 433–434, 2009.

[75] R. B. Hayes, Y. Songnian, M. Dosemeci, and M. Linet, "Benzene and lymphohematopoietic malignancies in humans," *American Journal of Industrial Medicine*, vol. 40, no. 2, pp. 117–126, 2001.

[76] T. A. McDonald, *Public Health Goal for Benzene in Drinking Water*, Office of Environmental Health Hazard Assessment, Sacramento, Calif, USA, 2001, http://oehha.ca.gov/water/phg/pdf/BenzeneFinPHG.pdf.

[77] W. J. Blot, L. A. Brinton, J. F. Fraumeni, and B. J. Stone, "Cancer mortality in U.S. counties with petroleum industries," *Science*, vol. 198, no. 4312, pp. 51–53, 1977.

[78] R. G. Olin, A. Ahlbom, and I. Lindberg-Navier, "Occupational factors associated with astrocytomas: a case-control study," *American Journal of Industrial Medicine*, vol. 11, no. 6, pp. 615–625, 1987.

[79] R. A. Lyons, S. P. Monaghan, M. Heaven, B. N. C. Littlepage, T. J. Vincent, and G. J. Draper, "Incidence of leukaemia and lymphoma in young people in the vicinity of the petrochemical plant at Baglan Bay, South Wales, 1974 to 1991," *Occupational and Environmental Medicine*, vol. 52, no. 4, pp. 225–228, 1995.

[80] J. Kaldor, J. A. Harris, and E. Glazer, "Statistical association between cancer incidence and major-cause mortality, and estimated residential exposure to air emissions from petroleum and chemical plants," *Environmental Health Perspectives*, vol. 54, pp. 319–332, 1983.

[81] M. Gérin, J. Siemiatycki, M. Désy, and D. Krewski, "Associations between several sites of cancer and occupational exposure to benzene, toluene, xylene, and styrene: results of a case-control study in Montreal," *American Journal of Industrial Medicine*, vol. 34, no. 2, pp. 144–156, 1998.

[82] J. D. Everall and P. M. Dowd, "Influence of environmental factors excluding ultra violet radiation on the incidence of skin cancer," *Bulletin du Cancer*, vol. 65, no. 3, pp. 241–247, 1978.

[83] P. Boffetta, N. Jourenkova, and P. Gustavsson, "Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons," *Cancer Causes and Control*, vol. 8, no. 3, pp. 444–472, 1997.

[84] M. S. Gottlieb, C. L. Shear, and D. B. Seale, "Lung cancer mortality and residential proximity to industry," *Environmental Health Perspectives*, vol. 45, pp. 157–164, 1982.

[85] C. Y. Yang, M. F. Cheng, J. F. Chiu, and S. S. Tsai, "Female lung cancer and petrochemical air pollution in Taiwan," *Archives of Environmental Health*, vol. 54, no. 3, pp. 180–185, 1999.

[86] B. J. Pan, Y. J. Hong, G. C. Chang, M. T. Wang, F. F. Cinkotai, and Y. C. Ko, "Excess cancer mortality among children and adolescents in residential districts polluted by petrochemical manufacturing plants in Taiwan," *Journal of Toxicology and Environmental Health*, vol. 43, no. 1, pp. 117–129, 1994.

[87] A. K. Hurtig and M. San Sebastián, "Geographical differences in cancer incidence in the Amazon basin of Ecuador in relation to residence near oil fields," *International Journal of Epidemiology*, vol. 31, no. 5, pp. 1021–1027, 2002.

[88] C. Paz-y-Miño, B. Castro, A. López-Cortés et al., "Impacto genético en comunidades amazónicas del Ecuador localizadas en zonas petroleras," *Revista Ecuatoriana de Medicina y Ciencias Biológicas*, vol. 1, no. 1-2, pp. 7–19, 2010.

[89] G. Thodi, F. Fostira, R. Sandaltzopoulos et al., "Screening of the DNA mismatch repair genes MLH1, MSH2 and MSH6 in a Greek cohort of Lynch syndrome suspected families," *BMC Cancer*, vol. 10, article 544, 2010.

[90] D. A. Lawes, T. Pearson, S. SenGupta, and P. B. Boulos, "The role of *MLH1, MSH2* and *MSH6* in the development of multiple colorectal cancers," *British Journal of Cancer*, vol. 93, no. 4, pp. 472–477, 2005.

[91] Environment Agency, *Principles for Evaluating the Human Health Risks from Petroleum Hydrocarbons in Soils: A Consultation Paper*, Environment Agency, Bristol, UK, 2003, http://www.environment-agency.gov.uk/static/documents/Research/petroleum_hydrocarbons1.pdf.

[92] I. Kovalchuk, O. Kovalchuk, and B. Hohn, "Biomonitoring the genotoxicity of environmental factors with transgenic plants," *Trends in Plant Science*, vol. 6, no. 7, pp. 306–310, 2001.

[93] C. Paz-y-Miño, A. Creus, O. Cabré, and P. E. Leone, *Genética, Toxicología y Carcinogenesis*, PUCE, Quito, Ecuador, 2002.

[94] T. K. Hei, R. Persaud, H. Zhou, and M. Suzuki, "Genotoxicity in the eyes of bystander cells," *Mutation Research*, vol. 568, no. 1, pp. 111–120, 2004.

[95] A. V. Carrano, "Chromosome aberrations and radiation-induced cell death. I. Transmission and survival parameters of aberrations," *Mutation Research*, vol. 17, no. 3, pp. 341–353, 1973.

[96] E. De Souza and J. P. D. M. Soares, "Occupational and technichal correlations of interventional radiology," *Jornal Vascular Brasileiro*, vol. 7, no. 4, pp. 341–350, 2008.

[97] J. A. V. Butler, "Effects of ultra-violet light on nucleic acid and nucleoproteins and other biological systems," *Experientia*, vol. 11, no. 8, pp. 289–293, 1955.

[98] D. R. Boreham, "Cellular defense mechanisms against the biological effects of ionizing radiation," in *Proceedings of the 10th International Congress of International Radiation Protection Association (IRPA '00)*, Hiroshima, Japan, May 2000, http://w3.tue.nl/fileadmin/sbd/Documenten/IRPA_refresher_courses/Cellular_Defense_Mechanisms_Against_the_Biological_Effects_of_Ionizing_Radiation.pdf.

[99] J. Chung, H. Ward, K. Teschke, P. A. Ratner, and Y. Chow, *A Retrospective Cohort Study of Cancer Risks among Nurses in British Columbia: Potential Exposure to Ionizing Radiation Report*, British Columbia: Research Secretariat of the Workers' Compensation Board of British Columbia, 2005, http://www.cher.ubc.ca/PDFs/Ionizing_Radiation_2005.pdf.

[100] G. Obe, P. Pfeiffer, J. R. K. Savage et al., "Chromosomal aberrations: formation, identification and distribution," *Mutation Research*, vol. 504, no. 1-2, pp. 17–36, 2002.

[101] V. Garaj-Vrhovac and D. Zeljezic, "Comet assay in the assessment of the human genome damage induced by γ-radiation in vitro," *Radiology and Oncology*, vol. 38, no. 1, pp. 43–47, 2004.

[102] C. Paz-y-Mino, P. E. Leone, M. Chavez et al., "Follow up study of chromosome aberrations in lymphocytes in hospital workers occupationally exposed to low levels of ionizing radiation," *Mutation Research*, vol. 335, no. 3, pp. 245–251, 1995.

[103] A. V. Carrano, "Chromosome aberrations and radiation-induced cell death. II. Predicted and observed cell survival," *Mutation Research*, vol. 17, no. 3, pp. 355–366, 1973.

[104] A. P. Krishnaja and N. K. Sharma, "Transmission of γ-ray-induced unstable chromosomal aberrations through successive mitotic divisions in human lymphocytes in vitro," *Mutagenesis*, vol. 19, no. 4, pp. 299–305, 2004.

[105] P. K. Gadhia, M. Gadhia, S. Georje, K. R. Vinod, and M. Pithawala, "Induction of chromosomal aberrations in mitotic chromosomes of fish *Boleophthalmus dussumieri* after exposure in vivo to antineoplastics Bleomycin, Mitomycin-C and Doxorubicin," *Indian Journal of Science and Technology*, vol. 1, no. 1, pp. 1–6, 2008.

[106] R. S. Cardoso, S. Takahashi-Hyodo, P. Peitl Jr., T. Ghilardi-Neto, and E. T. Sakamoto-Hojo, "Evaluation of chromosomal aberrations, micronuclei, and sister chromatid exchanges in hospital workers chronically exposed to ionizing radiation," *Teratogenesis Carcinogenesis and Mutagenesis*, vol. 21, no. 6, pp. 431–439, 2001.

[107] L. Hlatky, R. K. Sachs, M. Vazquez, and M. N. Cornforth, "Radiation-induced chromosome aberrations: insights gained from biophysical modeling," *BioEssays*, vol. 24, no. 8, pp. 714–723, 2002.

[108] R. Gertler, R. Rosenberg, D. Stricker et al., "Telomere length and human telomerase reverse transcriptase expression as

markers for progression and prognosis of colorectal carcinoma," *Journal of Clinical Oncology*, vol. 22, no. 10, pp. 1807–1814, 2004.

[109] K. I. Nakamura, E. Furugori, Y. Esaki et al., "Correlation of telomere lengths in normal and cancers tissue in the large bowel," *Cancer Letters*, vol. 158, no. 2, pp. 179–184, 2000.

[110] V. Balachandar, B. L. Kumar, K. Suresh, and K. Sasikala, "Evaluation of chromosome aberrations in subjects exposed to environmental tobacco smoke in Tamilnadu, India," *Bulletin of Environmental Contamination and Toxicology*, vol. 81, no. 3, pp. 270–276, 2008.

[111] M. S. Sierra-Torres, Y. Y. Arboleda-Moreno, L. S. Hoyos, and C. H. Sierra-Torres, "Chromosome aberrations among cigarette smokers in Colombia," *Mutation Research*, vol. 562, no. 1-2, pp. 67–75, 2004.

[112] C. Paz-y-Miño, J. C. Pérez, V. Dávalos, M. E. Sánchez, and P. E. Leone, "Telomeric associations in cigarette smokers exposed to low levels of X-rays," *Mutation Research*, vol. 490, no. 1, pp. 77–80, 2001.

[113] K. B. S. Kumar, R. Ankathil, and K. S. Devi, "Chromosomal aberrations induced by methyl parathion in human peripheral lymphocytes of alcoholics and smokers," *Human and Experimental Toxicology*, vol. 12, no. 4, pp. 285–288, 1993.

[114] S. D. Bouffler, M. A. Blasco, R. Cox, and P. J. Smith, "Telomeric sequences, radiation sensitivity and genomic instability," *International Journal of Radiation Biology*, vol. 77, no. 10, pp. 995–1005, 2001.

[115] Y. Saitoh, Y. Harata, F. Mizuhashi, M. Nakajima, and N. Miwa, "Biological safety of neutral-pH hydrogen-enriched electrolyzed water upon mutagenicity, genotoxicity and subchronic oral toxicity," *Toxicology and Industrial Health*, vol. 26, no. 4, pp. 203–216, 2010.

[116] U. Von Recklinghausen, C. Johannes, L. Riedel, and G. Obe, "Aberration patterns and cell cycle progression following exposure of lymphocytes to the alkylating agent Trenimon," *Chromosome Alterations*, pp. 315–324, 2007.

[117] L. C. Sánchez-Peña, B. E. Reyes, L. López-Carrillo et al., "Organophosphorous pesticide exposure alters sperm chromatin structure in Mexican agricultural workers," *Toxicology and Applied Pharmacology*, vol. 196, no. 1, pp. 108–113, 2004.

[118] J. Friedman, F. Shabtai, L. S. Levy, and M. Djaldetti, "Chromium chloride induces chromosomal aberrations in human lymphocytes via indirect action," *Mutation Research*, vol. 191, no. 3-4, pp. 207–210, 1987.

[119] P. L. Olive and J. P. Banáth, "The comet assay: a method to measure DNA damage in individual cells," *Nature Protocols*, vol. 1, no. 1, pp. 23–29, 2006.

[120] D. Milković, V. Garaj-Vrhovac, M. Ranogajec-Komor et al., "Primary DNA damage assessed with the comet assay and comparison to the absorbed dose of diagnostic X-rays in children," *International Journal of Toxicology*, vol. 28, no. 5, pp. 405–416, 2009.

[121] L. Jiunn-Wang, H. Ching-I, M. Isao, and C. Yng-Tay, "Chloroacetaldehyde indices chromosome aberrations and micronucleus formation but not 2-chloroethanol," *Journal of Health Science*, vol. 57, no. 3, pp. 300–303, 2011.

[122] L. C. Silva-Pereira, P. C. S. Cardoso, D. S. Leite et al., "Cytotoxicity and genotoxicity of low doses of mercury chloride and methylmercury chloride on human lymphocytes in vitro," *Brazilian Journal of Medical and Biological Research*, vol. 38, no. 6, pp. 901–907, 2005.

[123] T. Kyoya, Y. Obara, and A. Nakata, "Chromosomal aberrations in Japanese grass voles in and around an illegal dumpsite at the Aomori-Iwate prefectural boundary," *Zoological Science*, vol. 25, no. 3, pp. 307–312, 2008.

[124] K. Guleria and V. Sambyal, "Spectrum of chromosomal aberrations in peripheral blood lymphocytes of gastrointestinal tract (GIT) and breast cancer patients," *International Journal of Human Genetics*, vol. 10, no. 1–3, pp. 147–158, 2010.

[125] E. Horváthová, D. Slameňová, L. Hlinčíková, T. K. Mandal, A. Gábelová, and A. R. Collins, "The nature and origin of DNA single-strand breaks determined with the comet assay," *Mutation Research*, vol. 409, no. 3, pp. 163–171, 1998.

[126] H. Samavat and H. Mozdarani, "Chromosomal aberrations in Iranian radiation workers due to chronic exposure of X-irradiation," *International Journal of Low Radiation*, vol. 1, no. 2, pp. 216–222, 2004.

[127] J. M. Kauffman, "Radiation Hormesis: demonstrated, deconstructed, denied, dismissed, and some implications for public policy," *Journal of Scientific Exploration*, vol. 17, no. 3, pp. 389–407, 2003.

[128] M. J. Muñoz, A. López-Cortés, I Sarmiento, C. Herrera, M. E. Sánchez, and C. Paz-y-Miño, "Genetic biomonitoring of individuals exposed to ionizing radiation and the relationship with cáncer," *Oncología*, vol. 18, no. 1, pp. 75–82, 2008.

[129] N. Bayo, "Reacción celular ante la radiación," *Radiobiología*, vol. 1, no. 1, pp. 9–11, 2001.

[130] B. Leffon, B. Perez-Candahía, J. Loueiro, J. Mendez, and E. Pásaro, "Papel de los polimorfismos para enzimas de reparación en el daño del ADN inducido por estierno y estireno-7, 8-óxido," *Reviews in Toxicology*, no. 21, pp. 92–97, 2004.

[131] E. L. Goode, C. M. Ulrich, and J. D. Potter, "Polymorphisms in DNA repair genes and associations with cancer risk," *Cancer Epidemiology Biomarkers and Prevention*, vol. 11, no. 12, pp. 1513–1530, 2002.

[132] M. Díaz-Valecillos, J. Fernández, A. Rojas, J. Valecillos, and J. Cañizales, "Chromosome alterations in workers exposed to ionizing radiation," *Investigacion Clinica*, vol. 45, no. 3, pp. 197–211, 2004.

[133] F. Zakeri and T. Hirobe, "A cytogenetic approach to the effects of low levels of ionizing radiations on occupationally exposed individuals," *European Journal of Radiology*, vol. 73, no. 1, pp. 191–195, 2010.

[134] H. E. Jiliang, C. Weilin, J. Lifen, and J. Haiyan, "Comet assay and cytokinesis-blocked micronucleus test for monitoring the genotoxic effects of X-ray radiation in humans," *Chinese Medical Journal*, vol. 113, no. 10, pp. 911–914, 2000.

[135] L. Hagmar, S. Bonassi, U. Strömberg et al., "Chromosomal aberrations in lymphocytes predict human cancer: a report from the European study group on cytogenetic biomarkers and health (ESCH)," *Cancer Research*, vol. 58, no. 18, pp. 4117–4121, 1998.

[136] R. J. Albertini, D. Anderson, G. R. Douglas et al., "IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans," *Mutation Research*, vol. 463, no. 2, pp. 111–172, 2000.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 256063, 6 pages
doi:10.4061/2011/256063

Review Article

Relationship between DNA Mismatch Repair Deficiency and Endometrial Cancer

Kenta Masuda, Kouji Banno, Megumi Yanokura, Yusuke Kobayashi, Iori Kisu, Arisa Ueki, Asuka Ono, Nana Asahara, Hiroyuki Nomura, Akira Hirasawa, Nobuyuki Susumu, and Daisuke Aoki

Department of Obstetrics and Gynecology, Keio University School of Medicine, Shinanomachi 35 Shinjuku-Ku, Tokyo 160-8582, Japan

Correspondence should be addressed to Kouji Banno, kbanno@sc.itc.keio.ac.jp

Received 28 January 2011; Revised 25 August 2011; Accepted 20 September 2011

Academic Editor: Mark Berneburg

Some cases of endometrial cancer are associated with a familial tumor and are referred to as hereditary nonpolyposis colorectal cancer (HNPCC or Lynch syndrome). Lynch syndrome is thought to be induced by germline mutation of the DNA mismatch repair (MMR) gene. An aberration in the MMR gene prevents accurate repair of base mismatches produced during DNA replication. This phenomenon can lead to an increased frequency of errors in target genes involved in carcinogenesis, resulting in cancerization of the cell. On the other hand, aberrant DNA methylation is thought to play a key role in sporadic endometrial carcinogenesis. Hypermethylation of unmethylated CpG islands in the promoter regions of cancer-related genes associated with DNA repair leads to the cell becoming cancerous. Thus, both genetic and epigenetic changes are intricately involved in the process through which cells become cancerous. In this review, we introduce the latest findings on the DNA mismatch repair pathway in endometrial cancer.

1. Introduction

The incidence of endometrial cancer among malignant gynecological tumors has increased with lifestyle and environmental changes. In the US, 40,000 patients are diagnosed with endometrial cancer annually, and 7,500 patients die of this disease [1].

The number and prevalence of cases of endometrial cancer have increased worldwide and control of this cancer is urgently required. However, many aspects of the mechanism of carcinogenesis and pattern of advancement are unclear. Environmental factors such as obesity and a high estrogen level are thought to play important carcinogenic roles, but a close association with hereditary disposition has also been suggested, since double cancer and an increased incidence of cancer in relatives are common in patients with endometrial cancer.

Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), is a hereditary disease in which there is frequent development of colorectal, endometrial, and ovarian cancers. The cause is thought to be mutation of the DNA mismatch repair (MMR) gene in germ cells. However, the conventional explanation of the mechanism involving genetic changes—mutations of cancer-related genes—is inadequate and epigenetic changes in endometrial cancer are now being examined. In particular, aberrant DNA methylation is thought to play a key role in endometrial carcinogenesis. Breakdown of the DNA mismatch repair mechanism due to DNA hypermethylation plays a particularly important role in the development of endometrial cancer.

2. Lynch Syndrome

Lynch syndrome is a hereditary disease that includes frequent development of colorectal, endometrial, and ovarian cancers, and which is inherited in an autosomal dominant manner. Lynch syndrome is caused by a hereditary defect in the DNA mismatch repair (MMR) gene and the incidences in colorectal and endometrial cancers are 2-3% and 1-2%, respectively [2]. This syndrome was initially reported by

Table 1: Clinical diagnostic criteria for HNPCC.

Amsterdam minimum criteria (1990)
(1) At least 3 cases of colorectal cancer in relatives (verified pathologically)
(2) One is a first degree relative of the other two
(3) At least two successive generations should be affected
(4) One case of colorectal cancer diagnosed before the age of 50 years old
(5) FAP should be excluded
Revised amsterdam criteria II (1998)
(1) At least 3 relatives with an HNPCC-associated cancer (cancer of the colorectum, endometrium, small bowel, ureter or renal pelvis)
(2–5) As for the minimum criteria

Wartin et al. in 1913 in a family with a high risk of development of colorectal cancer. Subsequent analysis of this family led Lynch to propose the disease concept of cancer family syndrome in 1971 [3, 4].

Six variants of the MMR gene, the causative gene in Lynch syndrome, have been cloned: *hMSH2, hMLH1, hMSH3, hMSH6, hPMS1,* and *hPMS2*. An aberration in one of these genes prevents accurate repair of base mismatches produced during DNA replication. In Lynch syndrome with an *hMLH1* or *hMSH2* mutation, the frequencies of colorectal and endometrial cancers are 68% and 62%, respectively, and the lifetime risk of developing endometrial cancer is higher than that for colorectal cancer in women [5].

Diagnosis of Lynch syndrome is based on clinical criteria. In 1990, the International Collaborative Group- (ICG-) HNPCC established the following diagnostic criteria for HNPCC, which are referred to as the classical Amsterdam criteria: (1) HNPCC is diagnosed when 3 or more patients with histologically confirmed colorectal cancer are present in a family line and one is a first relative of the other two; (2) colorectal cancer develops over two generations; (3) one case is diagnosed at younger than 50 years old [6]. In 1999, the new Amsterdam Criteria (Amsterdam II) [7] (Table 1) were published. These criteria address endometrial cancer, small intestinal cancer, urethral cancer, and kidney cancer, in addition to the colorectal cancer included in the classic criteria.

3. DNA Mismatch Repair Gene and Endometrial Cancer

DNA mismatch repair (MMR) system corrects DNA base pairing errors in newly replicated DNA. Mispaired nucleotides may be present after DNA replication, along with small insertion/deletion mutations that tend to occur at repetitive sequences. The MMR system is an excision/resynthesis system that can be divided into 4 phases: (i) recognition of a mismatch, (ii) recruitment of repair enzymes, (iii) excision of the incorrect sequence, and (iv) resynthesis by DNA polymerase using the parental strand as a template. This system is conserved through evolution from bacteria to human [8].

An aberration in one of MMR genes prevents accurate repair of base mismatches produced during DNA replication, resulting in production of a DNA chain of altered length, particularly in highly repeated sequences (microsatellites). This phenomenon is called microsatellite instability (MSI) and can lead to an increased frequency of errors in target genes involved in carcinogenesis, resulting in cancerization of the cell. Among the MMR genes, germline mutations of *hMLH1* on chromosome 3 and *hMSH2* on chromosome 2 are thought to cause most cases of HNPCC. Mutation of *hMSH6* has also been proposed to be important for development of HNPCC-associated endometrial cancer, but the details are unclear (Figure 1) [9].

4. Clinical Diagnostic Criteria for Lynch Syndrome

Since the Amsterdam Criteria for Lynch syndrome were proposed in 1991, several other diagnostic criteria, including the Japanese Criteria and the Bethesda Criteria, have been published. The confusion caused by the different criteria was resolved by revision of the Amsterdam Criteria by the ICG-HNPCC in 1999, to give the new Amsterdam Criteria [7] (Table 1). These criteria address endometrial cancer, small intestinal cancer, urethral cancer, kidney cancer, and colorectal cancer. Cases not meeting the classical Amsterdam Criteria may meet the new Amsterdam Criteria, and this has resulted in an increased number of cases diagnosed as Lynch syndrome. In addition, discovery of Lynch syndrome is now possible through investigation of familial histories of endometrial cancer patients. The revision also recognized the importance of cooperation among gynecologists for identification of Lynch syndrome. However, one concern with the new criteria is the omission of ovarian, breast, and stomach cancer, which may also be associated with Lynch syndrome.

The 1999 revised Amsterdam criteria II include endometrial cancer as a Lynch syndrome-related tumor, but women who develop endometrial cancer as the initial cancer and patients with a family tree with insufficient details are not included; thus, a high false negative rate has been reported based on these criteria [5]. For colorectal cancer, the Bethesda criteria require MSI testing, but this is not applicable for patients who develop endometrial cancer as the initial cancer. Thus, there is a need to establish criteria for selection of patients with endometrial cancer who should undergo screening [10].

5. Carcinoma of the Lower Uterine Segment and Lynch Syndrome

Endometrial cancer arises from the uterine body and fundus in many cases, but can also originate from the lower region of the uterine body through the upper region of the cervix. Such tumors are referred to as carcinoma of the lower uterine segment (LUS) or isthmus and account for 3–6.3% of all cases of endometrial cancer. The association of carcinoma of the LUS with Lynch syndrome has attracted recent attention. The frequency of Lynch syndrome in general endometrial cancer is 1-2% [2]. In contrast, Lynch syndrome has a high

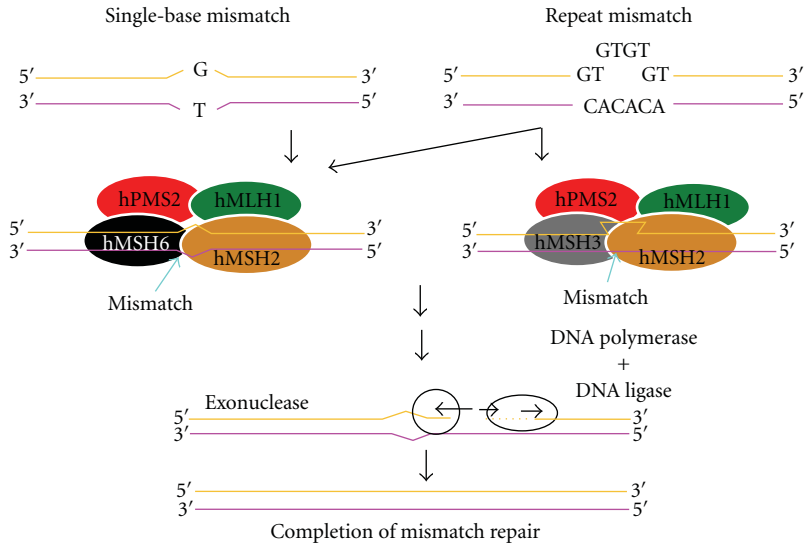


Figure 1

frequency in cases of carcinoma of the LUS, with one report in the US suggesting that 29% of such cases could also be diagnosed with Lynch syndrome and that the hMSH2 mutation was present at a high frequency in these cases [11]. Demonstration of an association between carcinoma of the LUS and Lynch syndrome in a large-scale survey would allow patients with carcinoma of the LUS to be classified as a high-risk group for Lynch syndrome [12].

6. Microsatellite Instability and Endometrial Cancer

Microsatellite instability occurs when the mismatch repair system is damaged. Microsatellites are DNA sequences of repeating units of 1 to 5 base pairs. Abnormalities in the mismatch repair system may cause replication errors in the repeating unit, leading to changes in length that are referred to as MSI. MSI caused by MMR gene aberration is detectable by PCR using microsatellite markers. In screening for Lynch syndrome, use of 5 microsatellite markers, two mononucleotide repeats (BAT26 and BAT25) and three dinucleotide repeats (D5S346, D2S123, and D17S250), is recommended [13]. MSI is observed in certain types of cancer, including 20 to 30% of cases of endometrial cancer [14]. These results suggest that MMR gene abnormalities occur frequently in endometrial cancer.

To investigate the status and characteristics of familial endometrial cancer, Banno et al. [15] surveyed the familial and medical histories of 385 patients who underwent treatment for endometrial cancer. MSI analysis was performed in 38 of these patients. The familial histories showed that 2 of the 385 cases met the new Amsterdam Criteria for Lynch syndrome, giving a rate of Lynch syndrome of about 0.5%. Investigation of familial accumulation of cancer in 890 relatives (439 men and 451 women) of the 38 endometrial cancer patients who underwent MSI analysis revealed high incidences of endometrial cancer, colorectal cancer, and ovarian cancer, suggesting that a hereditary factor common to Lynch syndrome is also involved in endometrial cancer. MSI analysis detected at least one of 5 microsatellite markers (D2S123, D3S1284, D5S404, D9S162, and hMSH2 intron 12) in 12 of the 38 cases (31.6%). This rate is very high compared to MSI in cancers of other organs, demonstrating that abnormal DNA mismatch repair plays an important role in endometrial cancer. The patients with MSI showed a tendency to have double cancer (such as ovarian cancer) compared with patients with microsatellite stability (MSS), although the difference was not significant (27% versus 15%). Regarding prognosis, none of the MSI-positive cases were fatal (0/11, 0%), while 5 MSI-negative (MSS) cases were fatal (5/27, 19%). The difference was not significant, but this tendency is similar to that for Lynch syndrome-associated colorectal cancer. The incidences of moderately differentiated adenocarcinoma G2 (36%) and poorly differentiated adenocarcinoma G3 (18%) tended to be higher in MSI-positive endometrial cancer, although again the difference was not significant. These findings appear contradictory with the favorable prognosis, but interestingly they may reflect the biological characteristics of endometrial cancer induced by abnormal DNA mismatch repair [16].

7. Screening for Endometrial Cancer and Prophylactic Hysterectomy in Lynch Syndrome

Women with Lynch syndrome have a high risk for endometrial cancer, with a life-long incidence of 40% to 60%, which is similar to or greater than that of colon cancer [17]. Therefore, a woman diagnosed with Lynch syndrome should undergo screening or prophylactic hysterectomy.

Potential screening methods include transvaginal ultrasound and endometrial biopsy. Transvaginal or transabdominal sonography is used to evaluate endometrial conditions and thickness. Some studies have shown a high false-positive

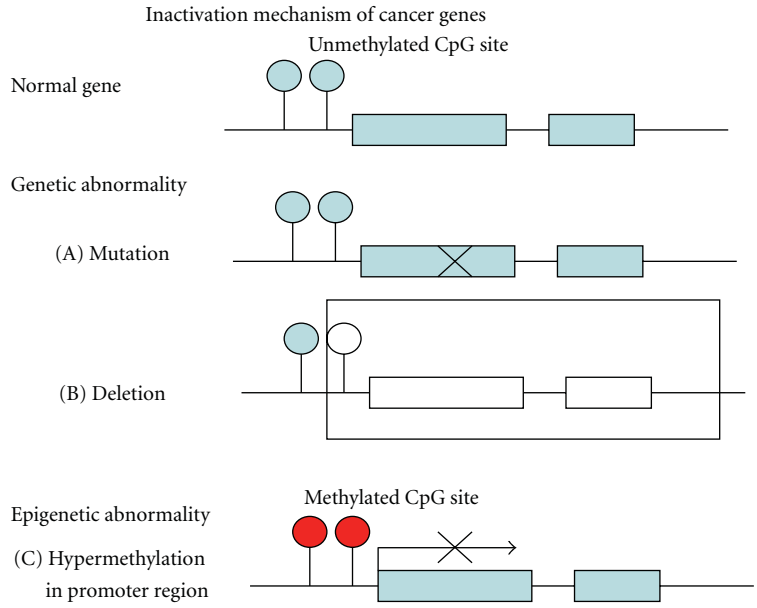


Figure 2

rate and poor efficacy [18, 19], while others have shown high sensitivity and negative predictive values [20]. Endometrial biopsy is not used for general screening but may be useful for patients with Lynch syndrome with a high risk for endometrial cancer. Thus, women who have a DNA mismatch repair gene mutation or a family history of this mutation should undergo a biopsy every year at the age of 30–35 [21].

One article has reviewed 5 papers reporting results of gynecological cancer surveillance in Lynch syndrome. Of the five articles included in this review, three were retrospective observational study. One study was prospective cohort study in Finland. Another study was a prospective clinical study, which evaluated the performance of hysteroscopy and endometrial biopsy in women at risk of Lynch syndrome in France. This article concluded that, although surveillance can detect premalignant lesions, it does not completely remove the risk of invasive cancer and it remains unclear whether surveillance for gynecological cancer in women with Lynch syndrome would significantly decrease mortality [22].

Prophylactic hysterectomy has not been thought to reduce the cancer risk in women with Lynch syndrome. In 1997, the Cancer Genetics Studies Consortium suggested that there was insufficient evidence to recommend that women with Lynch syndrome should have prophylactic surgery to reduce the risk of gynecologic cancer [23]. However, prophylactic hysterectomy has been realistically conducted in some institutions. The effects of prophylactic hysterectomy are of interest. Schmeler et al. [24] showed that prophylactic hysterectomy had a cancer-protective effect based on a retrospective cohort analysis in 315 women with a detected hMLH1, hMSH2, or hMSH6 germline mutation from 1973 to 2004. Outcomes were compared between 61 patients who underwent hysterectomy for prophylaxis or benign disease and 210 patients who did not undergo prophylactic hysterectomy. None of the 61 patients in the hysterectomy group developed endometrial cancer, whereas 69 (33%) in the nonhysterectomy group had endometrial cancer. These results indicate that prophylactic hysterectomy significantly decreased the development of endometrial cancer.

These results suggest that further studies should be conducted to compare the morbidity and mortality between screening using sonography or endometrial biopsy and prophylactic surgery.

8. DNA Hypermethylation and Endometrial Cancer

Epigenetics refers to the information stored after somatic cell division that is not contained within the DNA base sequence. Recent findings have shown that epigenetic changes—selective abnormalities in gene function that are not due to DNA base sequence abnormalities—play a significant role in carcinogenesis in various organs. In particular, the relationship between cancer and aberrant hypermethylation of specific genome regions has attracted attention. A completely new model for the mechanism of carcinogenesis has been proposed in which hypermethylation of unmethylated CpG islands in the promoter regions of cancer-related genes in normal cells silences these genes and leads to the cell becoming cancerous (Figure 2).

The main difference between epigenetic abnormalities and genetic abnormalities such as gene mutations is that epigenetic changes are reversible and do not involve changes in base sequence. This suggests that restoration of gene expression is possible and that epigenetic data may lead to important molecular targets for treatment. Attempts have begun to detect aberrant DNA methylation of cancer cells present in minute quantities in biological samples and to apply the results to cancer diagnosis, prediction of the risk of carcinogenesis, and definition of the properties of a

particular cancer. The MMR gene *hMLH1* is a typical gene that is silenced by DNA methylation. In endometrial cancer, *hMLH1* silencing is found in approximately 40% of cases and is an important step in the early stages of carcinogenesis, with the loss of DNA mismatch repair function proposed to lead to mutation of genes such as PTEN. In patients with endometrial cancer, Banno et al. found aberrant hypermethylation of *hMLH1*, APC, E-cadherin, and CHFR in 40.4%, 22.0%, 14.0%, and 13.3% of cases, respectively. A significant decrease in protein expression was found in patients with aberrant methylation of *hMLH1* ($P < 0.01$) and E-cadherin ($P < 0.05$), and aberrant methylation of *hMLH1* was also found in 14.3% of patients with atypical endometrial hyperplasia (AEH). However, no aberrant methylation of the four cancer-related genes was found in patients with a normal endometrium. These results indicate that aberrant methylation of specific genes associated with carcinogenesis in endometrial cancer does not occur in a normal endometrium. Aberrant methylation of *hMLH1* was most frequent, and the observation of this phenomenon in AEH, which is found in the first stage of endometrial cancer, supports the hypothesis that aberrant methylation of *hMLH1* is an important event in carcinogenesis in endometrial cancer [10, 25].

9. Conclusion

The DNA mismatch repair pathway is important in carcinogenesis of endometrial cancer. Recent analyses have shown that the MMR pathway can be impaired via both genetic and epigenetic mechanisms. Genetically, Lynch syndrome in cases of endometrial cancer is caused by a hereditary defect in the MMR gene. However, there have been fewer studies on endometrial cancer compared to colorectal cancer in patients with Lynch syndrome. Clarification of the pathology and development of screening and genetic tests are required for further progress in this area. Epigenetic research in endometrial cancer suggests that damage to the mismatch repair system plays a significant role in carcinogenesis and that DNA hypermethylation is important in this mechanism. Many attempts are currently being made to use epigenetic abnormalities as new methods of diagnosis and treatment based on control of methylation. Further studies of the genetic and epigenetic mechanisms may have potential for diagnosis, risk assessment, and treatment of endometrial cancer.

References

[1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 59, no. 4, pp. 225–249, 2009.

[2] H. Hampel, W. Frankel, J. Panescu et al., "Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients," *Cancer Research*, vol. 66, no. 15, pp. 7810–7817, 2006.

[3] H. T. Lynch and A. J. Krush, "The cancer family syndrome and cancer control," *Surgery Gynecology and Obstetrics*, vol. 132, no. 2, pp. 247–250, 1971.

[4] H. T. Lynch and J. F. Lynch, "Hereditary nonpolyposis colorectal cancer," *Seminars in Surgical Oncology*, vol. 18, no. 4, pp. 305–313, 2000.

[5] K. E. Resnick, H. Hampel, R. Fishel, and D. E. Cohn, "Current and emerging trends in Lynch syndrome identification in women with endometrial cancer," *Gynecologic Oncology*, vol. 114, no. 1, pp. 128–134, 2009.

[6] H. F. A. Vasen, J. P. Mecklin, P. M. Khan, and H. T. Lynch, "The international collaborative group on hereditary non-polyposis colorectal cancer," *Diseases of the Colon and Rectum*, vol. 34, no. 5, pp. 424–425, 1991.

[7] H. F. A. Vasen, P. Watson, J. P. Mecklin, and H. T. Lynch, "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome)proposed by the international collaborative group on HNPCC," *Gastroenterology*, vol. 116, no. 6, pp. 1453–1456, 1999.

[8] T. Jascur and C. R. Boland, "Structure and function of the components of the human DNA mismatch repair system," *International Journal of Cancer*, vol. 119, no. 9, pp. 2030–2035, 2006.

[9] K. Banno, M. Yanokura, Y. Kobayashi et al., "Endometrial cancer as a familial tumor: pathology and molecular carcinogenesis," *Current Genomics*, vol. 10, no. 2, pp. 127–132, 2009.

[10] Y. Muraki, K. Banno, M. Yanokura et al., "Epigenetic DNA hypermethylation: clinical applications in endometrial cancer," *Oncology Reports*, vol. 22, no. 5, pp. 967–972, 2009.

[11] S. N. Westin, R. A. Lacour, D. L. Urbauer et al., "Carcinoma of the lower uterine segment: a newly described association with Lynch syndrome," *Journal of Clinical Oncology*, vol. 26, no. 36, pp. 5965–5971, 2008.

[12] K. Masuda, K. Banno, M. Yanokura et al., "Carcinoma of the lower uterine segment (LUS): clinicopathological characteristics and association with Lynch syndrome," *Current Genomics*, vol. 12, no. 1, pp. 25–29, 2011.

[13] C. R. Boland, S. N. Thibodeau, S. R. Hamilton et al., "A national cancer institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer," *Cancer Research*, vol. 58, no. 22, pp. 5248–5257, 1998.

[14] T. Kanaya, S. Kyo, Y. Maida et al., "Frequent hypermethylation of MLH1 promoter in normal endometrium of patients with endometrial cancers," *Oncogene*, vol. 22, no. 15, pp. 2352–2360, 2003.

[15] K. Banno, N. Susumu, T. Hirao et al., "Two Japanese kindreds occurring endometrial cancer meeting new clinical criteria for hereditary non-polyposis colorectal cancer (HNPCC), Amsterdam criteria II," *Journal of Obstetrics and Gynaecology Research*, vol. 30, no. 4, pp. 287–292, 2004.

[16] K. Banno, N. Susumu, M. Yanokura et al., "Association of HNPCC and endometrial cancers," *International Journal of Clinical Oncology*, vol. 9, no. 4, pp. 262–269, 2004.

[17] M. Aarnio, R. Sankila, and H. J. Järvinen, "Cancer risk in mutation carriers of DNA-mismatch-repair genes," *International Journal of Cancer*, vol. 81, no. 2, pp. 214–218, 1999.

[18] F. E. M. Rijcken, M. J. E. Mourits, A. G. J. van der Zee, J. H. Kleibeuker, and H. Hollema, "Gynecologic screening in hereditary nonpolyposis colorectal cancer," *Gynecologic Oncology*, vol. 91, no. 1, pp. 74–80, 2003.

[19] I. Dove-Edwin, D. Boks, H. J. W. Thomas et al., "The outcome of endometrial carcinoma surveillance by ultrasound scan in women at risk of hereditary nonpolyposis colorectal carcinoma and familial colorectal carcinoma," *Cancer*, vol. 94, no. 6, pp. 1708–1712, 2002.

[20] F. Lécuru, C. Huchon, and P. L. Puig, "Contribution of ultrasonography to endometrial cancer screening in patients with hereditary nonpolyposis colorectal cancer/Lynch syndrome,"

International Journal of Gynecological Cancer, vol. 20, no. 4, pp. 583–587, 2010.

[21] N. M. Lindor, G. M. Petersen, N. Press et al., "Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review," *Journal of the American Medical Association*, vol. 296, no. 12, pp. 1507–1517, 2006.

[22] A. Auranen and T. Joutsiniemi, "A systemic review of gynecological cancer surveillance in women belonging to hereditary nonpolyposis colorectal cancer (Lynch syndrome) families.," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 90, no. 5, pp. 437–444, 2011.

[23] W. Burke, G. Petersen, C. Varricchio et al., "Recommendations for follow-up care of individuals with an inherited predisposition to cancer. I. Hereditary nonpolyposis colon cancer. Cancer genetics studies consortium," *Journal of the American Medical Association*, vol. 277, no. 11, pp. 915–919, 1997.

[24] K. M. Schmeler, H. T. Lynch, K. H. Lu et al., "Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome," *The New England Journal of Medicine*, vol. 354, no. 3, pp. 261–269, 2006.

[25] K. Banno, M. Yanokura, N. Susumu et al., "Relationship of the aberrant DNA hypermethylation of cancer-related genes with carcinogenesis of endometrial cancer," *Oncology reports*, vol. 16, no. 6, pp. 1189–1196, 2006.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 562849, 9 pages
doi:10.4061/2011/562849

Review Article

Priming DNA Replication from Triple Helix Oligonucleotides: Possible Threestranded DNA in DNA Polymerases

Patrick P. Lestienne

U 1053 INSERM, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux, France

Correspondence should be addressed to Patrick P. Lestienne, patrick.lestienne@inserm.fr

Received 14 April 2011; Accepted 2 July 2011

Academic Editor: Frédéric Coin

Triplex associate with a duplex DNA presenting the same polypurine or polypyrimidine-rich sequence in an antiparallel orientation. So far, triplex forming oligonucleotides (TFOs) are known to inhibit transcription, replication, and to induce mutations. A new property of TFO is reviewed here upon analysis of DNA breakpoint yielding DNA rearrangements; the synthesized sequence of the first direct repeat displays a skewed polypurine- rich sequence. This synthesized sequence can bind the second homologous duplex sequence through the formation of a triple helix, which is able to prime further DNA replication. In these case, the d(G)-rich Triple Helix Primers (THP) bind the homologous strand in a parallel manner, possibly via a RecA-like mechanism. This novel property is shared by all tested DNA polymerases: phage, retrovirus, bacteria, and human. These features may account for illegitimate initiation of replication upon single-strand breakage and annealing to a homologous sequence where priming may occur. Our experiments suggest that DNA polymerases can bind three instead of two polynucleotide strands in their catalytic centre.

1. Introduction

Since the pioneering work of Avery et al. is showing that DNA is the support of the genetic information [1] and the characterization of *E. coli* DNA polymerase I by A. Kornberg [2], DNA replication has been the matter of intense researches.

Even though, the description of DNA-dependent DNA polymerases seems to completion, the identification of new classes of enzymes, notably repair polymerases, has brought new concepts on DNA integrity, recombination (a major feature providing diversity), and thus on evolution. These proofreading activities may be downregulated leading to mutations and DNA rearrangements causing diseases.

Basically, replication involves DNA strand separation by DNA helicases, followed by the priming of complementary sequences, or by primase activities yielding Okazaki fragments on the lagging strand. DNA polymerases elongate the 3′ hydroxyl end of the primers on the template strand by phosphodiester bond, resulting from the enzymatic hydrolysis of complementary dNTP into dNMP and the release of pyrophosphate. The free energy resulting from this process is accompanied by a conformational change of the DNA polymerase from an opened to a closed conformation, together with the translocation of the double-stranded DNA.

Thus, DNA polymerases are known to bind only two DNA strands in their catalytic centre: the template strand and the complementary one under elongation.

In several cases of diseases, genetics bring the molecular basis of dysfunctions being related with pathology, yielding eventually new concepts in the biochemistry of molecular interactions, notably in DNA transactions mediated through enzymes, among which DNA and RNA polymerases, RNA processing, and their potentially modified properties.

Mitochondrial DNA, by its small size and thus information content, but indispensable one, may offer new insights into novel molecular mechanisms relevant to their downregulated structures; the description of which is thereby presented.

2. Skewed Base Composition at Mitochondrial DNA Rearrangement

Studies of mitochondrial DNA breakpoints inducing deletions and even partial tandem triplications, leading to diseases, were first described two decades ago [3, 4] and reviewed in [5]. They revealed that these DNA breakpoints mainly occur between two direct repeats (DRs) [6] separated

by several Kb of the 16,6 kb of the human mitochondrial DNA and were localized between the heavy and light strand origins of replication [7]. Analysis of the localization of these heteroplasmic deletions led to conclude that different mechanisms would occur since breakpoints appeared either at perfect repeats or at no repeat, or imperfect direct repeats [6, 8]. The first replicated DR1 was kept at the expense of DR2, and a slippage-mispairing model was proposed by Shoffner et al. involving a nuclease intermediate [9]. Perusal analysis of these DR, from 7 to 13 base pairs, revealed an unexpected information; the nucleotide composition of the synthesized DR1 is skewed in purine nucleotides (A, G) content [10]. This observation reminded us of the work of Felsenfeld et al. [11] who showed by titration that a poly(A) molecule could associate with two poly(U) molecules forming a triple helix. Such unexpected observation, done with a spectrophotometer 4 years after the discovery of the double helix structure, remained somehow on side until researchers found their potential and the general rule governing their formation ([12] for review). The third strand forming the triple helix has a similar sequence (either polypurine or polypyrimidine) than the homologous strand of the base-paired duplex, onto which it binds in an *antiparallel* orientation.

The observation that the synthesized DR1 is generally purine rich allowed us to propose that upon dissociation of the (neosynthesized DR1-DNA polymerase complex) and the further binding to the homologous DR2, the tertiary complex could prime and invade the duplex DR2 through the formation of the triple helix, which would be elongated on the double strand (Figure 1). A major issue was the *parallel* orientation required to the triple helix primer (THP) despite the known antiparallel binding of TFO [12, 13].

3. Designs of an Enzymatic Assay

To test this hypothesis, initial experiments by gel shift assays showed the association between a neosynthesized purine-rich DR1 of 10 nucleotides long with either a hairpin DNA presenting the binding site of 10-base pairs or the duplex DNA presenting the homologous strand in a parallel orientation. Then, the hairpin DNA containing the double-strand primer binding site was designed such that its 3′ end deoxyguanosine was substituted by a dideoxyguanosine, to prevent its own elongation; its 20 nucleotides-long 5′ end served as a single-strand template. The 5′ end of the potential triple helix primer (10 nucleotides long) was 5′ ^{32}P end labelled, so its elongation to 30 nucleotides would provide evidence for its priming property, which was indeed observed when tested with phage T7, T4, the Klenow Fragment, and to a lower extent by Taq and sequenase [14].

Nevertheless, according to the model, the THP had to be elongated on a double-strand that has not a current feature. To test this hypothesis, we used a similar system but with an entire double stranded DNA tethered by thymidine residues to maintain the strand stoichiometry. Results were negative [15]. Then A/A mismatches were introduced nearby the 3′ end of the 10-bases-long primer binding site. Results were positive using T7, Klenow Fragment, and Tth DNA polymerases [15]. Furthermore, a low elongation was evidenced with only one A/A mismatch located even 5 base pairs ahead of the 3′ end of the THP.

The unusual parallel orientation of the THP was further tested using an orthogonal method by annealing a polypurine strand to a polypyrimidine strand presenting an asymmetric primer binding site ending by a C/C mismatch to allow elongation in both orientations. Depending on the length of the replication product with restricted amounts of dNTP in the assay and the primer used and known to form triple helix [$d(A)_{10}$ or $d(T)_{10}$] (Figure 2), the elongated primer would yield a 50 nucleotides-long product if the lower strand is replicated with the primer $d(A)_{10}$ in a parallel orientation, while it would give a 30 nucleotides-long product or no elongation if the upper strand is the template strand, and thus if the primer is in an antiparallel orientation. Conversely, using $d(T)_{10}$ as a primer, parallel orientation would produce a 30-nucleotides-long fragment while an antiparallel orientation would yield a 50 nucleotides-long product or no elongation [16]. Results from Figure 2 showed a product of about 50 nucleotides-long and thus that the lower strand is replicated from $d(A)_{10}$, while the upper one is replicated with $d(T)_{10}$ as a primer since it provides a 30 nucleotides-long product. These data led us to conclude without ambiguity the parallel orientation of the THP despite their reported antiparallel binding. Other polynucleotides were tested and gave similar results, especially when the pyrimidine primer of 11-residues-long (CTTCTTTCTTC) was used and even at a pH above 8.

These results contrast sharply with previous data showing their *antiparallel* binding and their effect on the inhibition of replication on single strand [17], double strand [18], and even on RNA elongation [19].

4. Generalization of These Observations

Since these properties could be restricted to phage and bacterial DNA polymerases, to a single THP 5′d(TGGGGAGGGG)3′, and to binding site 5′d(20 nt..CCCCTCCCCA T4 TGGGGAGGGG..20 nt)3′. complementary nucleotides, but presenting generally A/A mismatches nearby the 3′ end of the primer (DR 64 A) [15], we modified the primer and its binding site of 10 base pairs into a still 8 d(G)-rich primer long with 2 d(A): d(GGGAGGAGGG) (Pur 10). We tested other DNA polymerases, including HIV-1 and AMV reverse transcriptases, human DNA polymerase β, γ and λ as well as Dpo4 from the archaebacteria Sulfolobus *Solfataricus* [20] together with other oligonucleotides known to form triple helices [$d(A)_{10}$, $d(T)_{10}$, $d(C)_{10}$] [16]].

As shown in Figure 3, the repair DNA polymerase β elongates all of the tested primers, with the exception of the nonproductive complex with the template: primer 5′d(TGGGGAGGGG)3′ shown in lane 1 (DR 64) but with the notable exception of polymerase γ, which, precisely, has to replicate this DR mitochondrial DNA sequence. Furthermore, DR 62C presenting a C/C mismatch is elongated by all tested polymerases. In addition, all polymerases elongate the Pur10 primer with Pur 62 as a template as shown in lane 2. Thus, primers and binding sites allow primer extension,

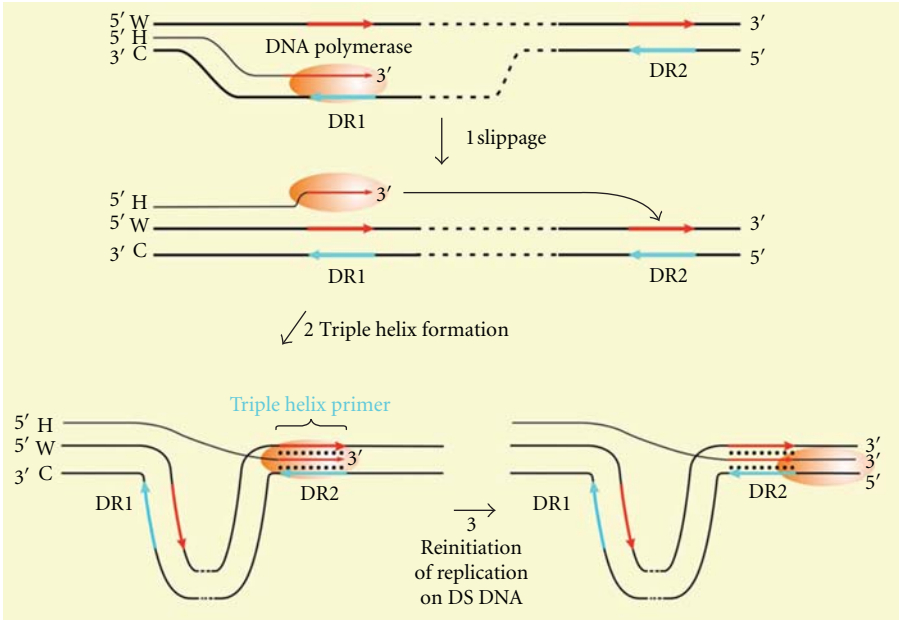


Figure 1: Model for the slippage mispairing between direct repeats. (1) Synthesized purine-rich DR1 sequence the [DNA polymerase-neosynthesized DNA complex] would dissociate from its template and bind to the homologous duplex sequence. (2) The triple helix bound to the duplex DNA is in a parallel orientation as the homologous strand and primes DNA replication on DR2. (3) Elongation of the THP on the DS DNA by DNA polymerase.

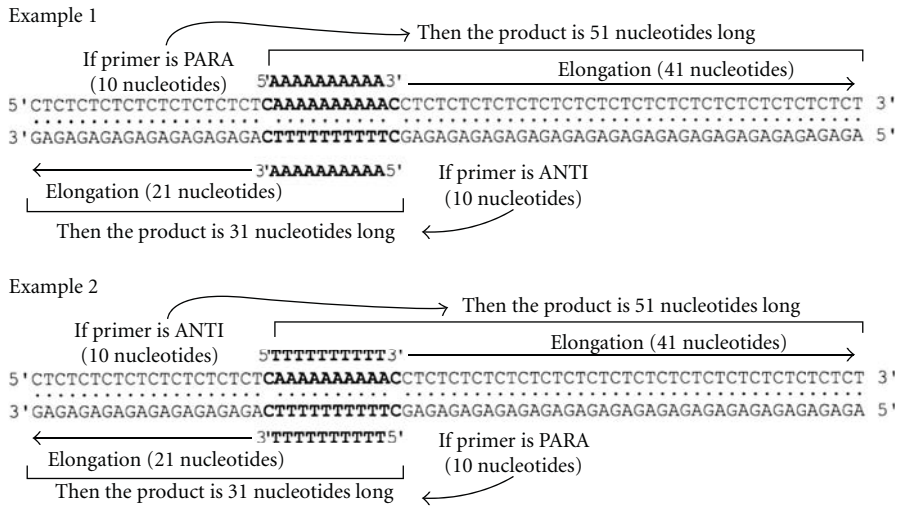


Figure 2: Determination of the orientation of the THP on a double-stranded DNA. The polypurine and polypyrimidine strands (50 pmol) were annealed at equimolar amounts. 5 pmoles of ^{32}P-end-labelled primers were tested with their corresponding binding site ending with a C/C mismatch in a replication assay with 5.10–8 M human DNA polymerase β. The Reaction media contained either 0.2 mM dGTP, and dATP or 0.2 mM dGTP, dCTP and dTTP. If the primer $(A)_{10}$ was parallel to the homologous strand, then, in the presence of dCTP and dTTP, there is synthesis of a 50 nt-long product. If the orientation was antiparallel, then a 30 nt-long product would be synthesized in the presence of dATP and dGTP. Conversely, if the $d(T)_{10}$ primer is parallel to its homologous primer binding site, then a product of 30 nt-long would be observed in the presence of dGTP, dCTP, and dTTP, while no product was observed with dATP and dGTP [16].

depending on their sequence. However, Dpo4 does not elongate $d(A)_{10}$, and HIV-1 and AMV reverse transcriptases, Pol γ, and Pol λ are unable to elongate $d(A)_{10}$, $d(T)_{10}$, and $d(C)_{10}$. Thus d(G)A-rich THP/PBS are the most common productive sequences [16]. These results rule out the trivial hypothesis of the opening of the hairpin templates, thus single stranded, followed by the binding of the complementary primer, as a duplex, otherwise all the results would have been similar. However, they indicate differences among DNA polymerases, likely relevant to their unwinding and catalytic properties. Kinetic studies with DNA Pol β showed comparable elongation rates for the primer Pur 10 and for a single- or double-stranded template [16]. Therefore, the sequence of the THP and its binding site significantly affect the efficient

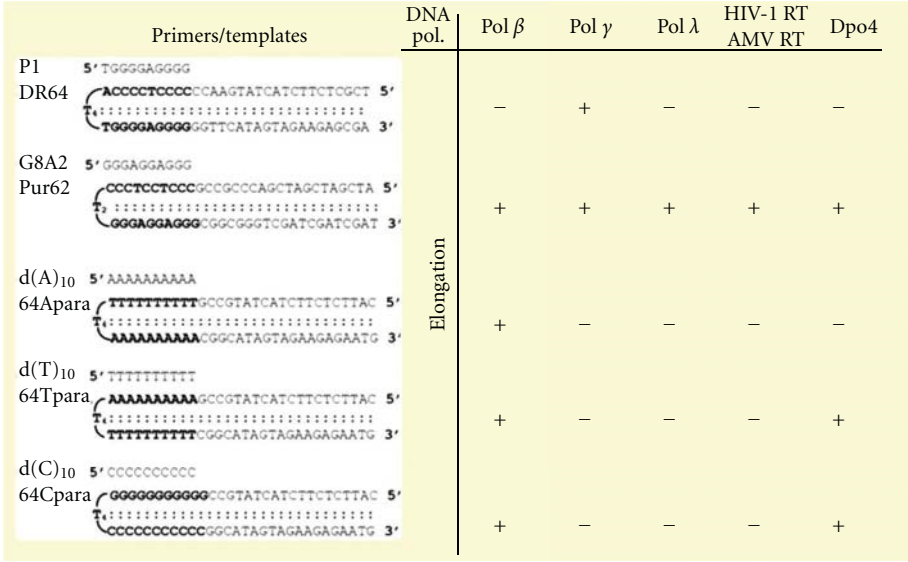


Figure 3: Results of elongation by several DNA polymerases with various templates/primers. Additional templates in an antiparallel orientation (i.e., 3′->5′ instead of 5′->3′) and presenting 3 A/A mismatches were nonproductive as well as 5′d(GGATTACGAG)3′ in a parallel and an antiparallel orientation.

elongation, even though primers P1 d(TGGGGAGGGG)3′ and Pur10 d(GGGAGGAGGG)3′ display 8 deoxyguanosine residues. This may be due to the peculiar structure of the P1-binding site due to two G4 tracts, as shown by NMR studies.

5. Microheterology May Be Tolerated between the Triple Helix Primer and Its Binding Site

To further test the slippage mispairing hypothesis, primer Pur10 length was elongated by 30 nucleotides at its 5′ end (Pur 40) to detect the specificity of its priming property. To this end, elongations on templates Pur 62 and DR 64 as a control were tested, while presenting, or not, a C/C mismatch at position 11 (Pur 62/Pur 62C) and (DR 64, DR 62C), at the theoretical 3′ end of the THP Pur 40. All the tested polymerases elongated Pur 40 with Pur 62 and Pur 62 C as templates in agreement with the previous data. As expected, the control DR 64 was not elongated by any polymerase but by polymerase γ. In contrast, using the heterologous template presenting a single C/C mismatch, DR 62C elongation of Pur 40 was evidenced with DNA polymerase β, λ, HIV-1, and AMV reverse transcriptases and Dpo4. Thus a single C/C mismatch enables elongation of the heterologous primer/primer binding site (PBS). A comparison between the THP and PBS is given on Figure 4. Mismatches of 3 to 4 nucleotides, over 10 nucleotides, of the THP /PBS may still enable DNA elongation; conversely, 6–8 d (G-A) residues at potential similar positions yield productive elongations [16]. These experiments indicate that mismatches between the THP and its PBS may account for rearrangements between imperfect repeats, as pointed before, since as few as 6-7 deoxyguanosine residues may prime DNA replication.

6. Biological Implications

During replication, the unwound lagging strand is exposed to DNA breakage by superoxide anions, which, by chance, may occur at polypurine-or polypyrimidine tracts when not protected by single-strand binding proteins. Thus, they could prime a partly homologous double-stranded DNA and invade it.

The d(G) content is nonrandom in eukaryote genomes, and many G tetrads promote deletions unless peculiar helicases unwind them. d(G) tracts are nonrandomly scattered within the genome but display a skewed representation in oncogene and tumor suppressor genes [21]. Furthermore, guanosine-rich telomeric sequences can stimulate DNA polymerase [22], and d(G)-rich tracts may be cleaved by endonuclease G, yielding 3′OH ends [23]. Triple helices may form either from polypurine/polypyrimidine sequences, H-DNA that are highly recombinogenic [24], together with intramolecular and palindromic structures prone to nuclease cleavage. Furthermore, an increasing number of diseases are associated with DNA forming triple helices [25] and rearrangements, including deletions, duplications, and even triplications, aside the example provided on mitochondrial DNA. The well-established translocation t(14;18) between Bcl2 and IgH inducing follicular lymphoma results from cleavage of one of the 3 dG tracts of the MBR (multiple break region) partly homologous to a sequence of IgH. This sequence forms an intramolecular triple helix with an unpaired strand [26], the complex of which may be cleaved, and the resulting rearrangements have been reported to be partly templated [27]. Z DNA- and non-B-structures are also nuclease sensitive and may be prone to cleavage or recombination [28]. This was even shown in the *E. coli Chi* sequence 5′d(GCTGGTGG) where exonuclease activity

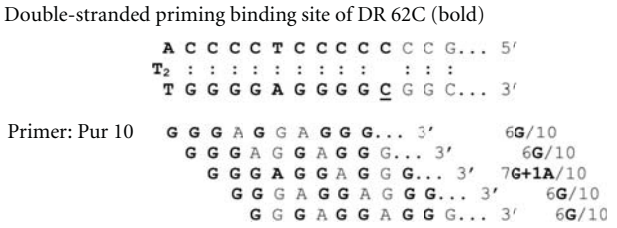


FIGURE 4: Microheterologies between the primer binding site of DR 62 C, and the 3′ end of primer Pur 40.

stops [29] and might play a role of primer. Furthermore, aside its d(G)-rich sequence, it is similar to the human VTRI, which is highly recombinogenic [30]. In addition, the promoter of the *c-myc* oncogene also displays a triple helix and a G tetrad [31] also found in the telomerase reverse transcriptase activity directed by the RNA template sequence [32]. Perfect or imperfect DR are involved in many cases of DNA rearrangements. Their sequence may slightly differ from one to another, an observation which may be related by tolerating the microheterogeneities shown above, and which may be as short as 6 base pairs long, a very close limit to those encountered in DNA rearrangements yielding diseases. Interestingly, a report showed that DNA invasion occurs before deletion in *Drosophila* [33].

7. Possible Shift between the Triple Helix Primer and Its Homologous Strand of the Duplex Binding Site: A RecA Analogy

The mechanisms of recombination involve the first step of synapsis implying recognition of a double strand by a homologous single strand. This reaction is mediated by RecA in *E. coli*, and Rad 51 in eukaryotes. The RecA nucleofilament recognises a homologous duplex sequence, thus forming a triplex with the homologous strands in a parallel orientation (for review see [34]). Then, the homologous strands are exchanged upon ATP hydrolysis. A structural organization of the base pairings between the double strand and a homologous strand of parallel orientation has been proposed by Zhurkin et al. [35]. Basically the third strand binds to the duplex through bonds with each nucleotide involved in Watson-Crick pairings. This model differs from the Hoogsteen bonds involving recognition of a nucleotide by the complementary one, while in an antiparallel orientation like the homologous nucleotide.

A schematic representation of the pairings described in [35] is shown on Figure 5. Interestingly, the third "parallel" polynucleotide is at about 90 degrees from the homologous one, instead of 180 degrees for the Hoogsteen bonds. The third-strand bound with the Watson-Crick one, in the model of Zhurkin that has been generalized, could account for our results (Figure 6). A recent three dimensional model of the RecA nucleofilament has given high information on this process where the single-strand bond to the RecA polymers displays a B structure able to exchange the DNA strand by stretching the double strand [36] upon the binding of ATP between each monomer.

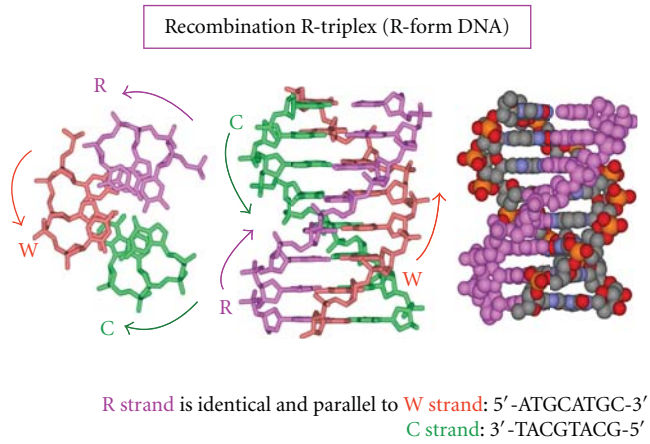


FIGURE 5: Molecular model of recombination according to Zhurkin et al. [35] (with kind permission). The R strand represents the triple-stranded intermediate in recombination strands mediated by RecA, or THP in our case, while the red molecule seen from the top is the Watson displaced and parallel strand. The middle of the figure shows the same but perpendicular triple helix. The right panel displays the third strand in the triplex conformation.

Moreover, a displacement of 45 degrees towards the homologous nucleotide yields displaced but still similar hydrogen bonds. A further rotation of 45 degrees would then displace the former Watson-Crick bonds. This mechanism requires energy, which, in DNA polymerase, could be brought by dNTP hydrolysis, formation of a phosphodiester bond with the 3′ end of the primer, and conformational change of the DNA polymerase.

8. DNA and RNA Polymerases: Catalytic and Structural Comparisons

DNA polymerase I displays a 3′-5′ proofreading and 5′-3′ exonuclease activities. Mild proteolysis yields the Klenow fragment devoid of 3′–5′ activity. The sequence of DNA polymerase I was published [37], its secondary structure predicted [38], and its tertiary structure determined [39]. The replicative phage T7 DNA polymerase was characterized with an unexpected processivity factor and thioredoxin [40] and their three dimensional structure with the template; primer DNA in the catalytic centre was determined [41]. Similarities between DNA polymerases were pointed from their sequences homologies [42] and by their catalytic centre shown by structural studies [43]. They may be compared to a right hand with a palm, a thumb, and fingers. Polymerisation of DNA occurs in the palm between the thumb and the finger by an acid amino acids triad (Asp, Asp, and Glu) and two magnesium ions [44].

Bound to the DNA polymerases, the DNA is generally in the B form but is condensed into the A form during elongation, at least in Taq polymerase [45]. The two bases at the 3′ end of the primer are in the A form with a helical twist and a larger minor groove [45]. A similar transition has been described for complexes with β polymerase [46, 47]. Contrasting with the two latter ones, template-primer DNA of T7

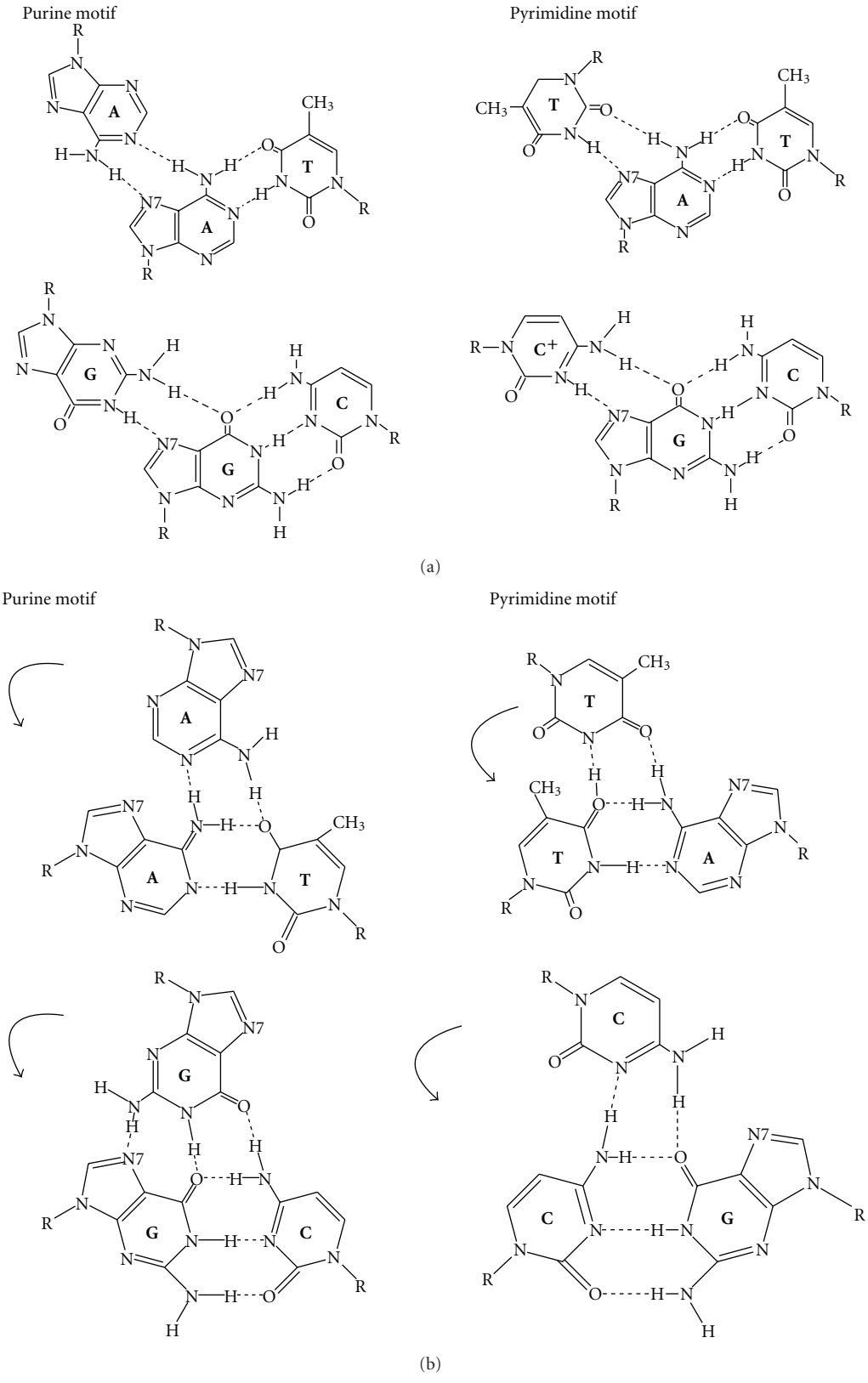


Figure 6: Model of interacting bonds with the R strand, which pairs with Watson-Crick bases modified from Zhurkin et al. [35]. (a) show the classical interactions between triple helix and DS DNA with Hoogsteen and reverse Hoogsteen bonds. Shown in (b) the models could account for the displacement of the homologous Watson strand by the primer or the R strand as in recombination with the invaded duplex without mismatches. The most stable triplexes are G: **G:C**, C: **C:G**, A: **A:T**, and T: **T:A**.

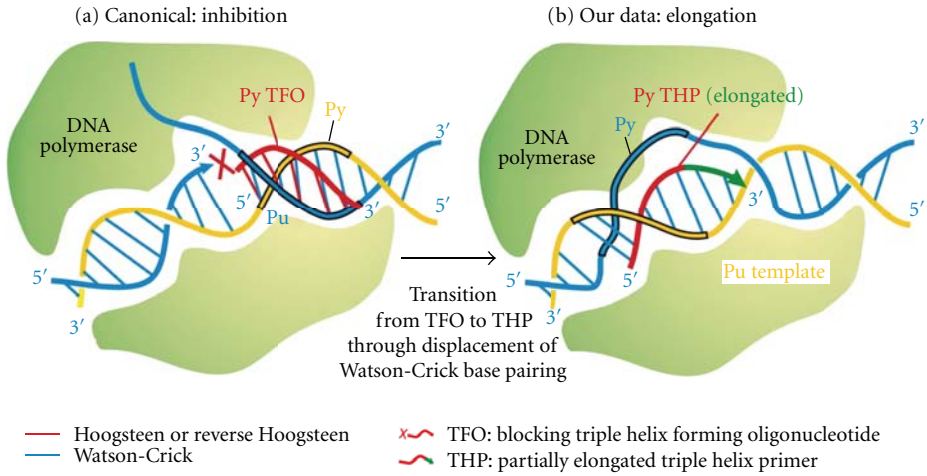


Figure 7: Comparison between the previously documented inhibition of polymerases by triple helices, and interpretation of our data.

DNA polymerase is curved into an S, formed by numerous interactions with the thumb and the fingers. Between 5 to 8 base pairs are away from the catalytic centre [44].

The 5′ end of the template binds the finger's surface. The template/primer is contacted by the finger at the catalytic centre by phosphodiester bonds. The 3′ end of the primer is anchored by the finger and the palm, and is bound to the thumb at its 5′ end. The thumb pushes against the minor groove at 5-6 bases pairs of the catalytic centre with two helices and a loop close to the top of the thumb. Amino acids in the contact of the template strand are localized at the bottom of the helices. DNA polymerase I strand displacement is favoured by the finger subdomains with amino acid residue Arg 84, which interacts with the template, and Ser 769, Phe 771, which favour strand separation [48].

Finally, DNA polymerase I discrimination between rNTP and dNTP occurs during the transition from the opened to the closed conformation of the enzyme [49, 50], while a mutation at amino acid 526 (Tyr → Phe) has been shown to be critical for the discrimination between NTP to ddNTP [51]. In the eukaryote DNA polymerase ι, Hoogsteen base pairings have been discussed [52].

The RNA polymerases activities, notably of phage T7 [53], are characterized by a pretranslocation [54] and an elongation phase [55]. Several rNMP are added, but the proofreading occurs during this step as RNA polymerase goes forward and reverse on the promoter region formed by an A/T-rich bubble. RNA polymerisation is similar as for DNA polymerases involving an acidic amino acid triad (2 Asp, 1 Glu) and 2 magnesium ions. As for DNA polymerase, the enzyme conformation differs during each nucleotide incorporation. The dissociation of the pyrophosphate upon hydrolysis of the rNTP enables a conformational change, strand separation, and translocation.

The unwinding of the promoter occurs between base pairs −17 to −5. From −4 to 1, there is the formation of the RNA heteroduplex [56] with 2 DNA stands and 1 RNA strand. The RNA: DNA hybrid is in the A form similarly to the last base pairs of the primer: template complex of DNA polymerase. Of interest, a phage T7 RNA polymerase mutant changing Tyr 639 into Phe enables the incorporation of dNTP like in DNA polymerases, reverse transcriptase, DNA- or RNA-directed polymerization depending on the template and the nucleotide in the assay [57]. The elongation phase enables the synthesized RNA to be dissociated from DNA through a tunnel crossing the RNA polymerase. From structural experiments,Steitz et al. concluded to a common mechanism for polynucleotide synthesis by DNA-dependent RNA and DNA polymerases [58] and a convergent catalysis [59].

Therefore, our results showing DNA elongation from a triple helix primer by DNA polymerases seem to meet three aspects of RNA and DNA polymerases, as well as of RecA. Actually the triple helix may be the easiest way to accommodate three polynucleotide strands, via a RecA-like activity of DNA polymerases, upon NTP hydrolysis. Of interest to note, the primer is parallel to the homologous strand and may, thereby, as for RecA, displace the homologous strand in the DNA polymerase catalytic centre, base pairs with the template strand, forming thus a transient D-loop with the previous complementary strand (Figure 7). Secondly, DNA polymerases can bind 3 strands, as our results show. However, the way the third strand is dissociated from the double-stranded template is likely to differ from the RNA polymerase tunnel, unless it has not been found. Sequencing experiments revealed a double incorporation of ddNMP facing the transition from purine pyrimidine purine, as if the polymerase or the primer stuttered during elongation [15] or that Tth sequenase incorporated ddNMP facing the nontemplate strand sequence; similarly, termination chain reaction with DNA polymerase β showed additional ddNMP incorporation while using a DNA primer, though using an RNA primer, and elongation was as expected [16]. This recalls us the initial pretranslocation complex in RNA polymerases, which stutters before full elongation phase. Finally, RNA and DNA polymerases display the same convergent catalysis mediated by 2 magnesium ions and an acid amino-acids triad. Eventually, computer modelling showed that the third triple helix primer is lying in the major groove of

the double-stranded DNA, with the DNA polymerase of phage T7 [14] and Dpo4 [16] catalytic residues close to the 3′ end of the primer.

Further structural experiments may shed lights on this novel property of triple helices and of DNA polymerases under study since five decades.

Abbreviations

THP: Triple helix primer
PBS: Priming binding site
DR: Direct repeat
Kb: Kilo base.

Acknowledgments

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM). The author wishes to acknowledge the Ligue Régionale Contre le Cancer (Comité Gironde), the Canceropôle Grand Sud-Ouest, and C. Cazaux, J. Bonnet, F. Boudsocq, E. Marza, J. Rosenbaum, and V. Zhurkin for stimulating discussions.

References

[1] O. T. Avery, C. M. MacLeod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III," *Molecular Medicine*, vol. 1, no. 4, pp. 344–365, 1944.

[2] N. Kresge, R. D. Simoni, and R. L. Hill, "Arthur Kornberg' discovery of DNA polymerase I," *The Journal of Biological Chemistry*, vol. 280, no. 49, article e46, 2005.

[3] P. Lestienne and G. Ponsot, "Kearns-Sayre syndrome with muscle mitochondrial DNA deletion," *Lancet*, vol. 1, no. 8590, p. 885, 1988.

[4] I. J. Holt, A. E. Harding, and J. A. Morgan-Hughes, "Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies," *Nature*, vol. 331, no. 6158, pp. 717–719, 1998.

[5] P. Lestienne, M. F. Bouzidi, I. Desguerre, and G. Ponsot, "Molecular basis of mitochondrial DNA diseases," in *Mitochondrial Diseases: Models and Methods*, P. Lestienne, Ed., pp. 33–58, Springer, Paris, France, 1999.

[6] I. Nelson, F. Degoul, B. Obermaier-Kusser et al., "Mapping of heteroplasmic mitochondrial DNA deletions in Kearns-Sayre syndrome," *Nucleic Acids Research*, vol. 17, no. 20, pp. 8117–8124, 1989.

[7] E. A. Schon, R. Rizzuto, C. T. Moraes, H. Nakase, M. Zeviani, and S. DiMauro, "A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA," *Nature*, vol. 244, no. 4902, pp. 346–349, 1989.

[8] F. Degoul, I. Nelson, S. Amselem et al., "Different mechanisms inferred from sequences of human mitochondrial DNA deletions in ocular myopathies," *Nucleic Acids Research*, vol. 19, no. 3, pp. 493–496, 1991.

[9] J. M. Shoffner, M. T. Lott, A. S. Voljavec, S. A. Soueidan, D. A. Costigan, and D. C. Wallace, "Spontaneous Kearns-Sayre/chronic external ophthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: a slip-replication model and metabolic therapy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 20, pp. 7952–7956, 1989.

[10] C. Rocher, T. Letellier, W. C. Copeland, and P. Lestienne, "Base composition at mtDNA boundaries suggests a DNA triple helix model for human mitochondrial DNA large-scale rearrangements," *Molecular Genetics and Metabolism*, vol. 76, no. 2, pp. 123–132, 2002.

[11] G. Felsenfeld, D. R. Davies, and A. Rich, "Formation of a three-stranded polynucleotide molecule," *Journal of the American Chemical Society*, vol. 79, no. 8, pp. 2023–2024, 1957.

[12] M. D. Frank-Kamenetskii and S. M. Mirkin, "Triplex DNA structures," *Annual Review of Biochemistry*, vol. 64, pp. 65–95, 1995.

[13] C. Giovannangeli and C. Hélène, "Progress in developments of Triplex-Based strategies," *Antisense and Nucleic Acid Drug Development*, vol. 7, no. 4, pp. 413–421, 1997.

[14] C. Rocher, R. Dalibart, T. Letellier, G. Précigoux, and P. Lestienne, "Initiation of DNA replication by DNA polymerases from primers forming a triple helix," *Nucleic Acids Research*, vol. 29, no. 16, pp. 3320–3326, 2001.

[15] P. Lestienne, P. Pourquier, and J. Bonnet, "Elongation of oligonucleotide primers forming a triple helix on double-stranded DNA templates by purified DNA polymerases," *Biochemical and Biophysical Research Communications*, vol. 311, no. 2, pp. 380–385, 2003.

[16] P. P. Lestienne, F. Boudsocq, and J. E. Bonnet, "Initiation of DNA replication by a third parallel DNA strand bound in a triple-helix manner leads to strand invasion," *Biochemistry*, vol. 47, no. 21, pp. 5689–5698, 2008.

[17] V. S. Mikhailovi and D. F. Bogenhagen, "Termination within oligo(dT) tracts in template DNA by DNA polymerase γ occurs with formation of a DNA triplex structure and is relieved by mitochondrial single-stranded DNA-binding protein," *Journal of Biological Chemistry*, vol. 271, no. 48, pp. 30774–30780, 1996.

[18] A. S. Krasilnikov, I. G. Panyutin, G. M. Samadashwily, R. Cox, Y. S. Lazurkin, and S. M. Mirkin, "Mechanisms of triplex-caused polymerization arrest," *Nucleic Acids Research*, vol. 25, no. 7, pp. 1339–1346, 1997.

[19] M. Faria, C. D. Wood, L. Perrouault et al., "Targeted inhibition of transcription elongation in cells mediated by triplex-forming oligonucleotides," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 8, pp. 3862–3867, 2000.

[20] H. Ling, F. Boudsocq, R. Woodgate, and W. Yang, "Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication," *Cell*, vol. 107, no. 1, pp. 91–102, 2001.

[21] J. L. Huppert and S. Balasubramanian, "Prevalence of quadruplexes in the human genome," *Nucleic Acids Research*, vol. 33, no. 9, pp. 2908–2916, 2005.

[22] J. Ying, R. K. Bradley, L. B. Jones et al., "Guanine-rich telomeric sequences stimulate DNA polymerase activity in vitro," *Biochemistry*, vol. 38, no. 50, pp. 16461–16468, 1999.

[23] P. Widlak, L. Y. Li, X. Wang, and W. T. Garrard, "Action of recombinant human apoptotic endonuclease G on naked DNA and chromatin substrates: cooperation with exonuclease and Dnase I," *Journal of Biological Chemistry*, vol. 276, no. 51, pp. 48404–48409, 2001.

[24] G. Wang and K. M. Vasquez, "Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 37, pp. 13448–13453, 2004.

[25] J. J. Bissler, "Triplex DNA and human disease," *Frontiers in Bioscience*, vol. 12, pp. 4536–4546, 2007.

[26] S. C. Raghavan, P. Chastain, J. S. Lee et al., "Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation," *Journal of Biological Chemistry*, vol. 280, no. 24, pp. 22749–22760, 2005.

[27] U. Jäger, S. Böcskör, T. Le et al., "Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation," *Blood*, vol. 95, no. 11, pp. 3520–3529, 2000.

[28] A. Bianchi, R. D. Wells, N. H. Heintz, and M. S. Caddle, "Sequences near the origin of replication of the DHFR locus of Chinese hamster ovary cells adopt left-handed Z-DNA and triplex structures," *Journal of Biological Chemistry*, vol. 265, no. 35, pp. 21789–21796, 1990.

[29] S. C. Kowalczykowski, "Initiation of genetic recombination and recombination-dependent replication," *Trends in Biochemical Sciences*, vol. 25, no. 4, pp. 156–165, 2000.

[30] W. P. Wahls, L. J. Wallace, and P. D. Moore, "Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells," *Cell*, vol. 60, no. 1, pp. 95–103, 1990.

[31] B. P. Belotserkovskii, E. De Silva, S. Tornaletti, G. Wang, K. M. Vasquez, and P. C. Hanawalt, "A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription," *Journal of Biological Chemistry*, vol. 282, no. 44, pp. 32433–32441, 2007.

[32] T. R. Cech, "Beginning to understand the end of the chromosome," *Cell*, vol. 116, no. 2, pp. 273–279, 2004.

[33] M. McVey, J. R. LaRocque, M. D. Adams, and J. J. Sekelsky, "Formation of deletions during double-strand break repair in *Drosophila* DmBlm mutants occurs after strand invasion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 44, pp. 15694–15699, 2004.

[34] S. C. Kowaczykowski, D. A. Dixon, S. D. Eggelston, W. N. Lauder, and W. M. Rehrauer, "Biochemistry of homologous recombination," *Microbiological Reviews*, vol. 58, pp. 401–465, 1994.

[35] V. B. Zhurkin, G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero, and R. L. Jernigan, "A parallel DNA triplex as a model for the intermediate in homologous recombination," *Journal of Molecular Biology*, vol. 239, no. 2, pp. 181–200, 1994.

[36] Z. Chen, H. Yang, and N. P. Pavletich, "Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures," *Nature*, vol. 453, no. 7194, pp. 489–494, 2008.

[37] C. M. Joyce, W. S. Kelley, and N. D. Grindley, "Nucleotide sequence of the *Escherichia coli* polA gene and primary structure of DNA polymerase I," *Journal of Biological Chemistry*, vol. 257, no. 4, pp. 1958–1964, 1982.

[38] W. E. Brown, K. H. Stump, and W. S. Kelley, "*Escherichia coli* DNA polymerase I. Sequence characterization and secondary structure prediction," *Journal of Biological Chemistry*, vol. 257, no. 4, pp. 1965–1972, 1982.

[39] L. S. Beese, V. Derbyshire, and T. A. Steitz, "Structure of DNA polymerase I Klenow fragment bound to duplex DNA," *Science*, vol. 260, no. 5106, pp. 352–355, 1993.

[40] D. F. Mark and C. C. Richardson, "Escherichia coli thioredoxin: a subunit of bacteriophage T7 DNA polymerase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 73, no. 3, pp. 780–784, 1976.

[41] S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger, "Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution," *Nature*, vol. 391, no. 6664, pp. 251–258, 1998.

[42] A. Bernad, A. Zaballos, M. Salas, and L. Blanco, "Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases," *EMBO Journal*, vol. 6, no. 13, pp. 4219–4225, 1987.

[43] M. Delarue, O. Poch, N. Tordo, D. Moras, and P. Argos, "An attempt to unify the structure of polymerases," *Protein Engineering*, vol. 3, no. 6, pp. 461–467, 1990.

[44] T. A. Steitz, "DNA polymerases: structural diversity and common mechanisms," *Journal of Biological Chemistry*, vol. 274, no. 25, pp. 17395–17398, 1999.

[45] S. H. Eom, J. Wang, and T. A. Steitz, "Structure of Taq polymerase with DNA at the polymerase active site," *Nature*, vol. 382, no. 6588, pp. 278–281, 1996.

[46] M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, and J. Kraut, "Crystal structure of rat DNA polymerase β: evidence for a common polymerase mechanism," *Science*, vol. 264, no. 5167, pp. 1930–1935, 1994.

[47] H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, and J. Kraut, "Structures of ternary complexes of rat DNA polymerase β, a DNA template- primer, and ddCTP," *Science*, vol. 264, no. 5167, pp. 1891–1903, 1994.

[48] K. Singh, A. Srivastava, S. S. Patel, and M. J. Modak, "Participation of the fingers subdomain of *Escherichia coli* DNA polymerase I in the strand displacement synthesis of DNA," *Journal of Biological Chemistry*, vol. 282, no. 14, pp. 10594–10604, 2007.

[49] S. Doublié, M. R. Sawaya, and T. Ellenberger, "An open and closed case for all polymerases," *Structure*, vol. 7, no. 2, pp. R31–R35, 1999.

[50] C. M. Joyce, O. Potapova, A. M. DeLucia, X. Huang, V. P. Basu, and N. D. F. Grindley, "Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity," *Biochemistry*, vol. 47, no. 23, pp. 6103–6116, 2008.

[51] S. Tabor and C. C. Richardson, "A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 14, pp. 6339–6343, 1995.

[52] J. Wang, "DNA polymerases: Hoogsteen base-pairing in DNA replication?" *Nature*, vol. 437, no. 7057, pp. E6–E7, 2005.

[53] G. M. T. Cheetham and T. A. Steitz, "Structure of a transcribing T7 RNA polymerase initiation complex," *Science*, vol. 286, no. 5448, pp. 2305–2309, 1999.

[54] G. M. T. Cheetham, D. Jeruzalmi, and T. A. Steltz, "Structural basis for initiation of transcription from an RNA polymerase-promoter complex," *Nature*, vol. 399, no. 6731, pp. 80–83, 1999.

[55] Y. W. Yin and T. A. Steitz, "Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase," *Science*, vol. 298, no. 5597, pp. 1387–1395, 2002.

[56] V. N. Karamychev, I. G. Panyutin, R. D. Neumann, and V. B. Zhurkin, "DNA and RNA folds in the transcription complex as evidenced by Iodine-125 radioprobing," *Journal of Biomolecular Structure & Dynamics*, vol. 1, pp. 155–167, 2000.

[57] R. Sousa and R. Padilla, "A mutant T7 RNA polymerase as a DNA polymerase," *EMBO Journal*, vol. 14, no. 18, pp. 4609–4621, 1995.

[58] T. A. Steitz, S. J. Smerdon, J. Jäger, and C. M. Joyce, "A unified polymerase mechanism for nonhomologous DNA and RNA polymerases," *Science*, vol. 266, no. 5193, pp. 2022–2025, 1994.

[59] T. A. Steitz, "Visualizing polynucleotide polymerase machines at work," *EMBO Journal*, vol. 25, no. 15, pp. 3458–3468, 2006.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 542795, 12 pages
doi:10.4061/2011/542795

Review Article

Nucleotide Excision Repair in *Caenorhabditis elegans*

Hannes Lans and Wim Vermeulen

Department of Genetics, Medical Genetics Center, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Correspondence should be addressed to Hannes Lans, w.lans@erasmusmc.nl

Received 2 March 2011; Accepted 18 June 2011

Academic Editor: Giuseppina Giglia-Mari

Nucleotide excision repair (NER) plays an essential role in many organisms across life domains to preserve and faithfully transmit DNA to the next generation. In humans, NER is essential to prevent DNA damage-induced mutation accumulation and cell death leading to cancer and aging. NER is a versatile DNA repair pathway that repairs many types of DNA damage which distort the DNA helix, such as those induced by solar UV light. A detailed molecular model of the NER pathway has emerged from *in vitro* and live cell experiments, particularly using model systems such as bacteria, yeast, and mammalian cell cultures. In recent years, the versatility of the nematode *C. elegans* to study DNA damage response (DDR) mechanisms including NER has become increasingly clear. In particular, *C. elegans* seems to be a convenient tool to study NER during the UV response *in vivo*, to analyze this process in the context of a developing and multicellular organism, and to perform genetic screening. Here, we will discuss current knowledge gained from the use of *C. elegans* to study NER and the response to UV-induced DNA damage.

1. DNA Damage Response Mechanisms

To preserve and faithfully transmit DNA to the next generation, cells are equipped with a variety of DNA repair pathways and associated DNA damage responses, collectively referred to as the DNA damage response (DDR). DNA is continuously damaged by environmental and metabolism-derived genotoxic agents. It is vital for cells and organisms to properly cope with DNA damage, because unrepaired damage can give rise to mutation and cell death. The importance of the DDR is illustrated by several human cancer prone and/or progeroid hereditary diseases, which are based on defects in the DDR. Over the last decades, a wealth of information on the molecular mechanism of different repair pathways has been gathered from detailed *in vitro* and live cell studies. Currently, this acquired knowledge is being used to develop therapeutic strategies to treat patients suffering from the consequences of unrepaired DNA damage, such as cancer and aging [1].

Damage is repaired by different DNA repair pathways depending on the type of DNA lesion, genomic location, and the cell cycle phase (for reviews see [2–4]). Lesions originating from different genotoxic sources can range from small base modifications to double-strand breaks. Small base modifications, such as oxidative lesions which do not substantially distort the double helix, are repaired by base excision repair (BER). BER removes single or several bases and repairs the gap by DNA synthesis. Bigger lesions to one strand of the DNA which substantially distort the DNA helix are repaired by nucleotide excision repair (NER). NER repairs lesions by cutting out a patch of the damaged DNA strand and filling in the gap by DNA synthesis (see below and Figure 1). More rigid lesions, which covalently crosslink both strands of the DNA, are repaired by interstrand crosslink (ICL) repair. Its precise repair mechanism is still poorly understood, but it involves several unique proteins of the Fanconi Anemia pathway and proteins that function in other repair pathways as well. Finally, DNA double-strand breaks (DSBs) are repaired by either homologous recombination (HR) or nonhomologous end-joining (NHEJ) or other alternative DSB repair pathways. HR is an error-free repair pathway and utilizes a homologous chromosome or sister chromatid, which is only present in late S- or G2-phase of the cell cycle, to repair damage. NHEJ is capable of rejoining broken DNA ends also in G1-phase and noncycling cells. However, due to processing of DNA ends prior to ligation, NHEJ is more error prone

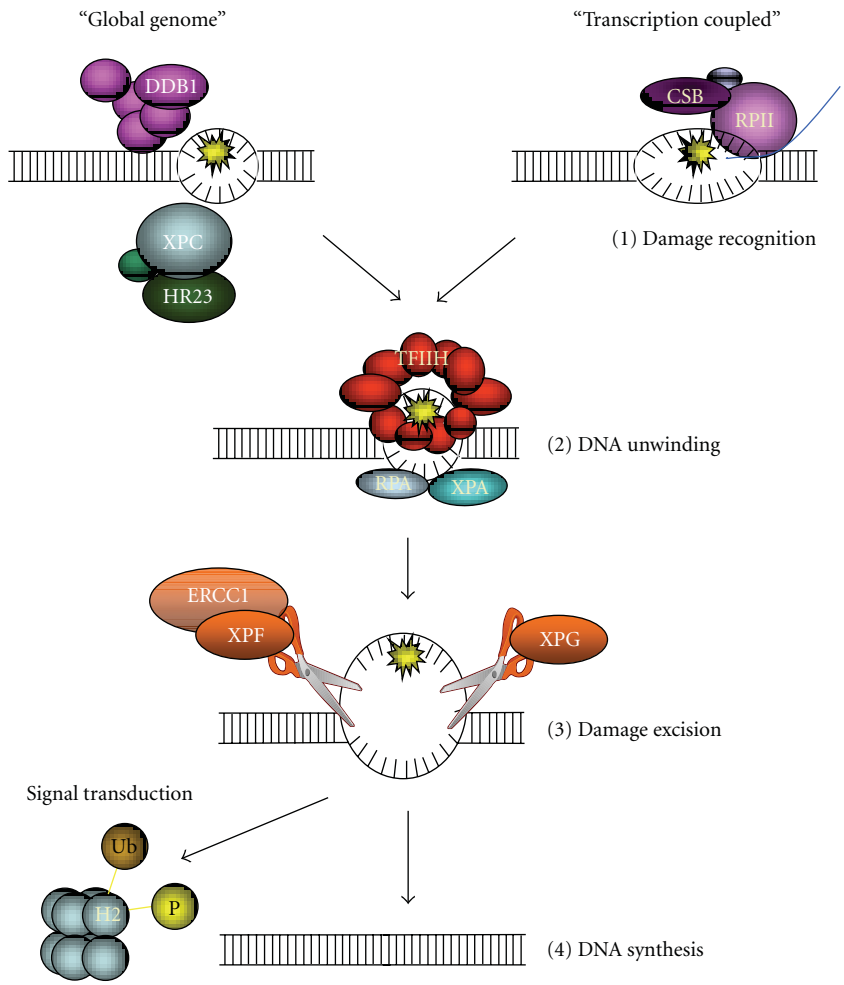


Figure 1: NER mechanism. DNA damage removal by NER is roughly executed in four subsequent steps. First, damage is recognised during transcription by stalling of RNA polymerase and involving CSB ("transcription coupled"), or it is recognised throughout the rest of the genome by the UV-DDB and XPC/HR23B complexes ("global genome"). Upon recognition, the TFIIH complex is recruited to unwind DNA around the damage and structural proteins XPA and RPA bind the resulting single-stranded DNA. Next, endonucleases ERCC1/XPF and XPG excise a patch of DNA including the damage. Finally, gap filling by *de novo* DNA synthesis takes place. During processing of a lesions, other proteins in proximity, including histones, are modified as part of a signalling cascade.

than HR. Although distinct classes of lesions are repaired by specific repair pathways, these pathways may compete for specific lesions or on the contrary, share common factors. In addition, several repair factors display multiple functions in DNA metabolism such as replication and transcription. These features show that the different repair pathways and other cellular responses to DNA damage form an interwoven intricate network. To fully understand DDR, it is, therefore, not sufficient to study a single repair process in isolation.

Much of the available knowledge regarding DDR mechanisms has come through the use of different model systems, such as bacteria, yeast, and cultured mammalian cells, and to a lesser extent of whole mice. The nematode *C. elegans* is increasingly being used to study various biological processes, including DNA repair [5–7]. This paper focuses on the function of NER in *C. elegans* and on the central role of this pathway in the cellular response to UV-induced DNA damage.

2. Nucleotide Excision Repair

Many organisms are continuously exposed to solar UV irradiation. Although the vast majority of UV light emitted by the sun is blocked by the earth's ozone layer, penetrating UV light can still severely damage DNA directly and is thought to be a major cause of skin cancer in humans [8]. UV radiation can cause a range of different DNA lesions of which cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (64PPs) are most abundant [9]. CPD and 64PP lesions distort the double helix structure of DNA, thereby severely impeding vital processes such as transcription and replication. If these lesions are not repaired properly, error-prone replication can induce mutations leading to cancer or cause cells to die, which contributes to aging [10]. Although predominantly skin cells of bigger organisms are exposed to solar UV irradiation, many other agents, such as chemotherapeutics, cigarette smoke, toxins, and some food-contained

chemicals, can cause similar helix distortions in other cells. Therefore, it is of major importance that cells are equipped with a mechanism to deal with this type of DNA damage. In bacteria, eukarya, and probably also in archaea, nucleotide excision repair (NER) functions to remove this wide range of helix-distorting lesions.

In the budding yeast *Saccharomyces cerevisiae* and in mammals, in which NER has been extensively studied, NER is executed in roughly four subsequent steps: (1) lesion detection, (2) local unwinding and damage verification, (3) incision of the DNA surrounding the lesion, and finally (4) DNA synthesis and ligation to fill the resulting gap (Figure 1; for reviews, see [11, 12]). DNA damage that occurs in the active strand of transcribed genes is repaired by transcription-coupled NER (TC-NER). TC-NER is thought to be initiated by stalling of RNA polymerase II on a lesion [13, 14] and involves recruitment of the ATP-dependent chromatin remodeling protein CSB and the WD40 domain containing protein CSA [15–17]. In yeast, the CSB ortholog Rad26 [18] is also implicated, but no functional homolog for CSA has been identified. DNA damage that occurs elsewhere in the genome is repaired by global genome NER (GG-NER). Such lesions are recognized in mammals by the UV-DDB ubiquitin ligase complex and the heterotrimeric complex XPC/hHR23/Centrin-2 [19–24]. In yeast, detection of lesions depends also on XPC and hHR23 orthologs Rad4 and Rad23 [25, 26], but it involves a different ubiquitin ligase complex consisting of Rad7, Rad16, Cul3, and Elc1 proteins [27]. Following detection of a lesion, the general transcription factor IIH (TFIIH) is recruited to the site of damage [28, 29]. Using its XPB and XPD helicase subunits, TFIIH locally unwinds a stretch of approximately 30 nucleotides around the lesion, providing access for other repair factors. Other essential DNA-binding proteins XPA and RPA are also recruited and thought to stimulate translocation and damage verification by TFIIH [30] and stabilize and orient the endonucleases which incise DNA around the damage [31]. Next, a stretch of approximately 25–30 nucleotides of DNA surrounding the lesion is excised by the structure specific endonucleases XPF/ERCC1 and XPG [32–34]. Finally, the resulting gap is filled in by DNA synthesis and ligation, involving replication factors such as PCNA and RFC and several ligases and polymerases [35–37].

The strong conservation of NER proteins across different life domains suggests that NER must be an important, universal repair pathway. This is also evident from the severe symptoms that are associated with NER deficiency in mammals [38]. Rare UV-sensitive hereditary disorders such as xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD) are caused by specific mutations in NER genes. XP is characterized by extreme cancer proneness, whereas CS and TTD exhibit segmental progeria and neurodevelopmental problems. In mice, these symptoms are phenocopied by similar mutations [39]. Some mutations in specific NER genes, such as those found in ERCC1 and XPF [40, 41], are associated with severe growth and developmental defects. This shows that even during normal growth, that is, even without excessive exposure to genotoxic agents, the function of these specific NER proteins is essential for normal development and life.

In spite of its strong evolutionary conservation, NER does not always function in exactly the same manner in different organisms. For instance, TC-NER does not seem to be active in *Drosophila* [42]. Furthermore, the yeast RAD23/RAD4 complex is not only involved in GG-NER, as in mammals, but also in TC-NER [43, 44]. Furthermore, cells from different tissues can respond differently to UV irradiation [45] or modify NER activity such that it is active in transcribed genes only [46]. It is thus likely that in different cells, other NER-regulatory pathways are activated. Although the basic mechanism by which NER removes and repairs damaged DNA is known, it is still not well understood to what level NER or individual NER proteins can be modulated or regulated and how this contributes to differential NER activity in cells. Besides NER, other mechanisms exist that process UV-induced DNA damage. In many organisms, photolyase enzymes directly reverse UV-induced DNA damage following activation by light [47]. This photoreactive repair is, however, not active in placental mammals. In proliferative cells, two emergency strategies can also prevent direct cell killing due to unrepaired UV lesions [48]. First, in S phase, UV lesions cause replication fork collapse and subsequent generation of DNA breaks, which are repaired by HR. Second, damage bypass mechanisms involving specialized translesion polymerases can circumvent damage in S-phase, in an error-prone way. However, to avoid these two unfavorable conditions, break induction, and low fidelity repair, most cells are equipped with efficient DNA damage signaling pathways that activate cell-cycle checkpoints providing more time to properly fix lesions.

3. NER in *C. elegans*

In recent years, the use of the nematode *C. elegans* to study DNA repair pathways has become increasingly intensive. *C. elegans*' main advantages for studying a biological processes such as DNA repair include its relatively fast and easy genetic manipulation, short life cycle, and straightforward recognizable *in vivo* phenotypes. The animal is simply grown on bacteria-seeded culture plates and produces self-fertilized offspring within a few days. Many loss-of-function mutants are available, and its genome, invariable cell lineage, and development are well annotated and accessible via various web resources (http://www.wormbase.org/). Homology searches, protein-protein interaction mapping analysis, and genetic screening have indicated that the major repair pathways found in mammals are conserved to the molecular level in *C. elegans* [6, 7, 49]. These pathways include BER [50–52], NHEJ, HR [5], ICL repair [7], mismatch repair [53, 54], and NER.

Almost 30 years ago the first UV-sensitive *C. elegans* mutants were identified and described [55]. These so-called *rad* (for abnormal radiation sensitivity, see Table 1) mutants were isolated in a screen for animals sensitive to UV or ionizing radiation. In subsequent years, phenotypes of these mutants were extensively characterized [55–61], but to date,

Table 1: List of original *rad* mutants.

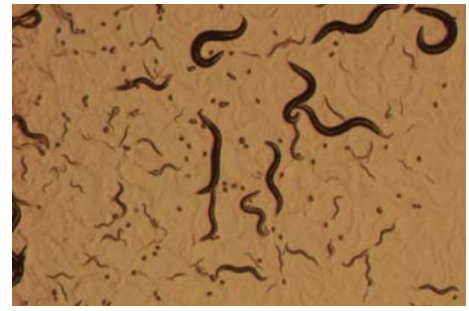
Locus	Gene	Sensitivity	Other affected processes	Repair	References
rad-1		UV, IR		Normal	[55, 56, 58]
rad-2		UV, IR, MMS	Embryonic checkpoint	Normal	[55, 56, 58, 64]
rad-3	*xpa-1*	UV, MMS	Germ line apoptosis	Absent	[55, 56, 58, 63, 65]
rad-4		UV, MMS	X chromosome nondisjunction		[55]
rad-5	*clk-2*	UV, IR	Checkpoint, longevity		[55, 62, 66]
rad-6		UV, IR	Embryogenesis		[55]
rad-7		UV, IR	Embryogenesis	Normal	[55, 56, 58]
rad-8		UV, oxygen	Embryogenesis, longevity		[55, 67]
rad-9		UV	Embryogenesis		[55]

the molecular identity of only two *rad* mutants is known. The *rad-5* locus encodes for the DNA damage checkpoint protein CLK-2 which is an ortholog of yeast Tel2p [62]. The *rad-3* locus encodes a genuine NER protein, the *C. elegans* XPA ortholog, which is essential for survival following UV irradiation [63].

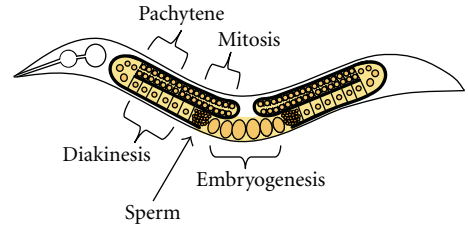
Orthologs of most of the known NER proteins have been identified by homology searches in *C. elegans* (Table 2). RNAi-mediated knockdown and loss-of-function mutations of many NER proteins has confirmed their role in the response to UV irradiation and repair of UV photolesions [63, 65, 68–73]. Importantly, no photolyases or photoreactive repair have been observed in *C. elegans* [58, 74]. Together, these results indicate that NER is fully operational and represents the major and only repair pathway which removes UV-induced DNA damage in *C. elegans*, just as in mammals. As will be discussed below, these studies also demonstrate the different response of different tissues to UV irradiation. Furthermore, NER is found to be differently regulated during development and aging and in cells of different tissues.

Because adult *C. elegans* consists of a limited set of 959 somatic cells which still represent many different cell types, the animal seems ideally suited to study the DDR *in vivo* during growth and development of different tissues. Adult *C. elegans* produce approximately 250–300 self-fertilized eggs, which hatch after a few hours. After hatching, larvae develop through four larval stages to become reproducing adults consisting of neuronal, muscular, epithelial, germ line, and other tissue (see Figure 2(a) for a mixed stage *C. elegans* culture). Dauer larvae represent a specialized developmental senescent "survival" stage, in between the L2 and L4 stage. Dauer larvae have been found to be more UV resistant than non-Dauer larvae [63], showing that the response to UV irradiation can change. Increased UV resistance may be a consequence of low levels of transcription in the dauer stage, such that damage does not interfere with this vital process, or because of Dauer-specific upregulation of prosurvival stress response pathways. Understanding how DNA damage leads to different responses in different cells might shed more light on the etiology of symptoms associated with human disease or cancer development.

In mammals, there is a clear distinction between DNA damage recognition via GG-NER and TC-NER, but in other organisms, both subpathways utilize the same proteins for damage recognition [11], or TC-NER may not function at all [42]. In *C. elegans*, UV-induced DNA damage in highly transcribed genes is more rapidly repaired than damage in poorly transcribed genes [65], which is in line with the existence of TC-NER. Furthermore, *C. elegans* expresses orthologs of the GG-NER specific XPC and hHR23B proteins, called XPC-1 and RAD-23, respectively, and an ortholog of the TC-NER specific CSB protein, called CSB-1. Epistatic analysis of mutant *rad-23*/*xpc-1* and *csb-1* animals suggests that these proteins act in parallel pathways in *C. elegans* [71]. Therefore, it is likely that also in *C. elegans* two separate DNA



(a) View of a *C. elegans* culture through the microscope. Worms of different stages can be discerned. Biggest worms are adults. Small dots are eggs. Worms of intermediate size are in different larval stages



(b) Schematic drawing of the two gonads and uterus of *C. elegans*. Only gonad structures, but not intestine and other tissues, are schematically depicted. Different areas in the gonads, containing germ cells in different developmental stages are indicated. Embryogenesis takes place in the uterus. Head is left, and tail is right

Figure 2: *C. elegans.*

Table 2: Nucleotide excision repair genes in *C. elegans*.

Mammalian gene		*C. elegans* gene	Available alleles	Sensitivity to UV	Additional affected processes	References
CETN2		*R08D7.5?*	*tm3611*			
DDB1		*ddb-1*	*tm1769*		Development, protein turnover	[75]
ERCC1		*ercc-1*	*tm1943*			
			tm1981			
			tm2073	yes		[71]
ERCC4 (XPF)		*xpf-1*	*e1487*	yes	Meiosis	[71, 76–79]
			tm2842	yes*		
			ok3039	no*		
ERCC5 (XPG)		*xpg-1*	*tm1670*	yes	Apoptosis	[71, 73]
			tm1682	yes	Apoptosis	[71, 73]
ERCC6 (CSB)		*csb-1*	*ok2335*	yes		[71]
LIG1		*lig-1*	RNAi		Growth, development	
LIG3		*K07C5.3*				
PCNA		*pcn-1*	*ok1905*		Growth, development	
			tm3157			
			tm3241		Growth, development	
POLD1		*F10C2.4*			embryogenesis	[80, 81]
POLH		*polh-1*	RNAi		DNA repair	[64, 79, 82]
POLK		*polk-1*	RNAi		Growth, embryogenesis, DNA repair	[79–81]
RAD23B		*rad-23*	*ok1910*			
			tm2595	yes		[71]
			tm3690			
RFC	RFC1	*rfc-1*	RNAi		Mutator, embryogenesis	[81, 83]
	RFC2	*rfc-2*	RNAi		Embryogenesis	[81]
	RFC3	*rfc-3*	RNAi		Mutator, embryogenesis	[81, 83]
	RFC4	*rfc-4*	RNAi		Growth, development	[80, 81]
	RFC5	*F44B9.8*	RNAI		Embryogenesis	
RPA	RPA1	*rpa-1*	RNAi		Embryogenesis	[80]
	RPA2	*rpa-2*	*ok1627*		Growth, development	
TFIIH	CCNH	*cyh-3*	RNAi		Growth, development	[80]
	CDK7	*cdk-7*	*ax224*		Transcription, cell cycle	[84]
	GTF2H1	*R02D3.3*	RNAi		Growth, development	[80]
	GTF2H2	*T16H12.4*	*tm1767*			
			tm4960			
	GTF2H3	*Zk1128.4*	*ok1200*		Growth, development	
			tm1501		Growth, development	
	GTF2H4	*Y73F8A.24*	RNAi		Embryogenesis	[81]
	GTF2H5	*Y55B1AL.2*				
	MNAT1	*mnat-1*	*tm2959*		Development	
	ERCC3 (XPB)	*Y66D12A.15*	RNAi		Apoptosis, embryogenesis	[73, 81, 85]
	ERCC2 (XPD)	*Y50D7A.2*			Apoptosis, embryogenesis	[73, 81, 85]
XPA		*xpa-1*	*mn157*	yes		[55, 63]
			ok698	yes	Apoptosis, lifespan?	[63, 65, 71, 86, 87]
			gk674			
XPC		*xpc-1*	*ok734*	no*	Apoptosis	[72]
			tm3886	yes	Apoptosis	[71, 72]

Some phenotypes were taken from http://www.wormbase.org/ (WS221). For those genes for which no alleles are known, results from RNAi experiments are indicated.

*Represent unpublished results.

damage recognition mechanisms, GG-NER and TC-NER, exist. However, despite this similarity to mammalian NER, *C. elegans* NER probably still functions slightly different as some NER proteins specifically implicated in GG-NER and TC-NER, DDB2 and CSA, have not yet been identified, and thus might not function in *C. elegans*.

4. UV Irradiation of *C. elegans*

It is unknown whether UV irradiation really represents a major source of DNA damage for *C. elegans* in its natural habitat [88]. Still, NER is highly conserved and required to survive exposure to UV irradiation. In nature, *C. elegans*, like other organisms not continuously exposed to solar irradiation, might be more likely to encounter genotoxic chemicals in its food and environment which induce DNA alterations that are targets for NER. However, to study NER, UV irradiation is often used as convenient tool to reproducibly and instantaneously induce large amounts of DNA lesions. Effective and reproducible UV irradiation experiments with the worm depend on several conditions. Most studies use UV-C light (254 nm) as damaging agent, which is very potent in generating 64PP and CPD photolesions, because it almost equals the maximum absorbance peak of DNA. An important drawback of UV-C light is its high absorption by water and biopolymers causing a low penetrance of tissue compared to UV light with higher wavelengths. Because *C. elegans* consists of multiple cell layers, we considered the application of a higher wavelength that still produces 64PP and CPD photolesions. Indeed, we have found that UV-B light (302 nm) produces a similar response of *C. elegans* as UV-C light, but it generates better reproducible results [71]. Another means to circumvent absorption problems might be the use of chemicals that induce lesions which are specifically processed by NER. However, such chemicals, like 4-nitroquinoline-1-oxide or N-acetyl-2-aminofluorene [89], have so far not been extensively tested in *C. elegans*.

When irradiating *C. elegans*, care must be taken with respect to shielding effects. For instance, the standard OP50 *E. coli* on which *C. elegans* is cultured [90] forms a relatively thick lawn that partially shields animals from UV light and causes variable results. To avoid this shielding, *C. elegans* should be irradiated in the absence of bacteria. Alternatively, if it is difficult to get rid of bacteria, HT115(DE3) *E. coli* can be used which form a thin lawn.

5. Repair Kinetics in *C. elegans*

It has been noted that late larval stages and adults (Figure 2(a)) are more resistant to UV radiation than younger animals, which can be partially attributed to shielding effects, as they are bigger in size. Shielding may also partially explain why multicellular organisms such as *C. elegans* can tolerate higher doses of UV irradiation than mammalian cells in monolayer culture. Size-related shielding effects are evident from the frequency of lesion induction by UV irradiation [58]. UV-C irradiation produces on average 0.4 to 0.5 lesions per 10 kb per 100 J/m^2 in young adult *C. elegans* [65]. In smaller animals such as L1 larvae, however, lesion frequency is higher, approximately 4 lesions/10 kb/100 J/m^2. Two studies have examined kinetics of UV-lesion repair, one using 64PP and CPD antibody-binding radioimmunoassay [58], the other using qPCR on the polymerase epsilon gene to detect polymerase-stalling lesions [65]. Both studies revealed that the global photolesion repair rate in *C. elegans* is comparable to that in cultured human cells but slower than in yeast and bacteria. However, in mammalian cells 64PPs are repaired at a much faster rate than CPDs [91], whereas in *C. elegans*, both photolesions are repaired at the same rate.

Initial repair rates immediately after UV irradiation seem to remain constant from embryogenesis to early adulthood although at later developmental stages more photoproducts remain unrepaired after 24 hrs [58]. In adulthood, repair seems to be biphasic in the sense that initial repair rates are higher than those after 24 hrs [65]. Furthermore, starting from adulthood, repair rates also decline as the animals age. This is not because global transcription levels or transcription of NER genes diminishes, but it was suggested to be correlated to reduced levels of ATP in aging animals. In general, NER genes are more highly expressed during embryogenesis than during adulthood. However, expression levels during adulthood remain constant and appear sufficient for repair to take place [65, 92].

6. UV Response of Germ Cells and Embryos

UV irradiation negatively influences germ cell and embryonic development, egg laying and male fertility [56, 71]. *C. elegans* germ cells are contained in two U-shaped gonads, which are joined together at their proximal ends to a common uterus (Figure 2(b); [93]). In the most distal parts of the gonads, germ nuclei mitotically proliferate and migrate in a proximal direction. Upon progression towards the uterus, nuclei further replicate DNA and enter meiosis prophase I. Just before the gonad bend (Figure 2(b)), meiotic nuclei exit pachytene stage, in which homologous chromosomes align and meiotic recombination is initiated and enter diplotene and subsequently diakinesis stage. Half of the germ cells in the pachytene stage are eliminated by apoptosis, probably to maintain tissue homeostasis [94]. As diakinesis stage cells further progress and pass through the spermatheca, they are fertilized, finish meiosis I and II, and initiate first cell divisions of embryogenesis. Because of its transparency and amenable manipulation, the *C. elegans* germ line has become an ideal tool to study DNA repair during the process of meiosis [6].

Progression and maturation of germ cells is blocked by DNA damage. Ionizing and UV radiation, as well as genotoxic chemicals, cause a transient cell cycle arrest of proliferating nuclei and increased apoptosis of pachytene stage nuclei [66]. Furthermore, unrepaired UV damage also blocks further maturation of pachytene cells and/or exit to diplotene [71]. Surprisingly, following UV irradiation, both induction of cell-cycle arrest and apoptosis depend on NER activity [72] and require either the GG-NER or TC-NER pathway [71]. Apoptosis induction furthermore

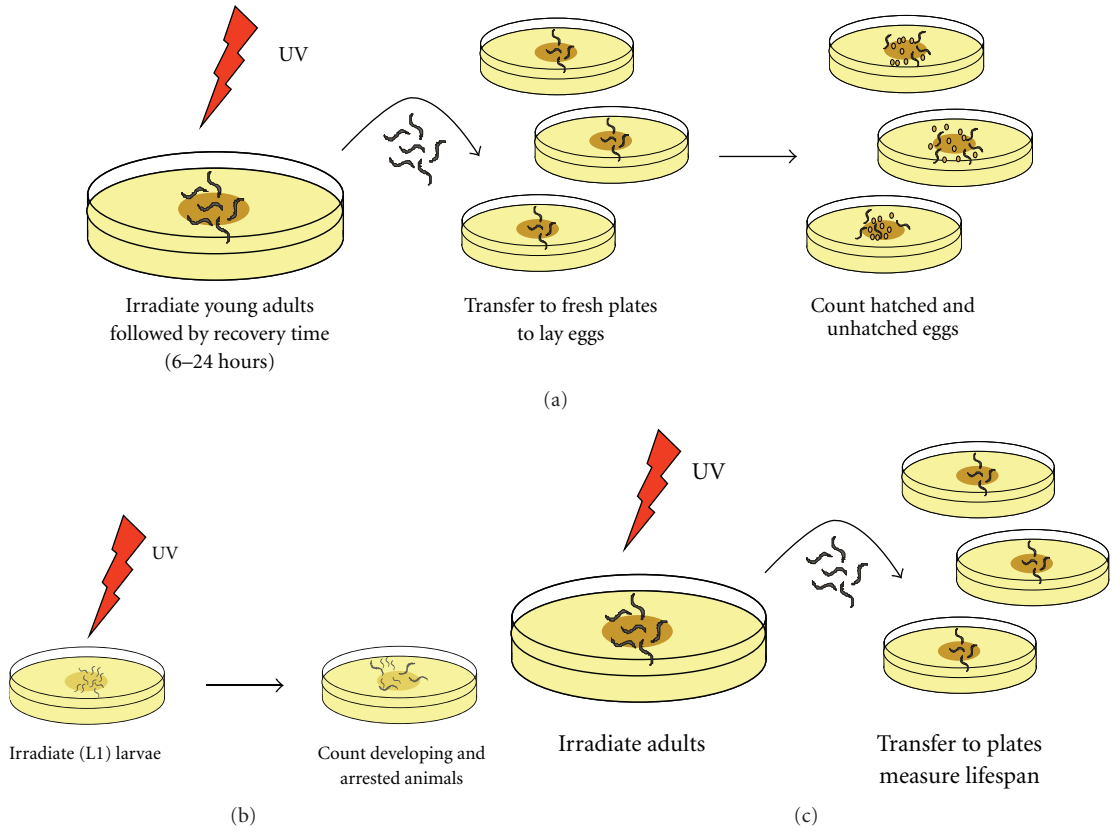


Figure 3: Assays to measure UV sensitivity. Shown are commonly used assays to measure survival of (a) germ cell and embryos, (b) larvae, and (c) adults.

involves the same checkpoint signaling proteins that are also involved after ionizing radiation, such as orthologs of the 9-1-1 signaling complex, the PI_3 kinases ATM and ATR and p53 [72]. Because UV-induced apoptosis and recruitment of checkpoint proteins also depends on some members of the HR pathway, such as orthologs of MRE11 and RAD54 but not RAD51, it was suggested that processing of a UV-lesion by NER is necessary for HR proteins to activate checkpoint signaling [73]. How this might be accomplished is still not understood, but it is reminiscent of studies in yeast and mammals which also show processing of UV lesions by NER as a prerequisite to activate checkpoint signaling [95, 96].

UV irradiation of the germ line causes embryonic lethality, especially if NER is compromised [63, 71, 72]. Embryonic lethality can be used as an easy readout to test whether a protein is involved in NER or the overall UV response (Figure 3(a)). Surprisingly, we found that survival of germ cells, meiotic maturation and repair of lesions after UV in the germ line specifically depends on GG-NER, as inactivation of only this pathway, and not TC-NER, renders germ cells sensitive to UV [71]. Only in a GG-NER deficient background does a TC-NER defect become essential, showing that TC-NER is active but that UV survival mainly depends on GG-NER. The importance of GG-NER in germ cells likely reflects the need of this immortal cell line to protect the integrity of the entire genome.

In contrast to germ cells, growth of *C. elegans* early embryos is relatively resistant to the induction of DNA damage by UV irradiation [55, 97]. Strikingly, a dose of UV which normally induces cell-cycle arrest in mitotic germ cells does not interfere with replication and timing of the cell cycle during the first embryonic cell division. This intriguing damage resistance was attributed to an actively suppressed checkpoint response, which otherwise halts the cell cycle for repair to take place [64]. Checkpoint suppression is dependent on the *rad-2* gene, one of the originally identified *rad* genes, and the *polh-1* and *gei-17* genes. The translesion polymerase POLH-1 is responsible for replicating damaged DNA. It is an ortholog of human POLH, which is mutated in the so-called variant form of the UV-sensitive XP syndrome [98]. In *C. elegans,* its knockdown also renders early embryos and germ cells hypersensitive to UV irradiation [82]. The PIAS1-related E3 SUMO ligase *gei-17* actively protects POLH-1 from being degraded after damage induction, thereby allowing replication of damaged DNA to occur [99].

NER proficient late embryos are even less sensitive to UV irradiation than early embryos [56, 71]. Paradoxically, late *xpa-1* mutant embryos seem to be more sensitive than early *xpa-1* embryos [56, 63]. It is not clear why this is, but it might involve differences in checkpoint silencing or transcription dependence.

7. Larval Development and Aging

UV irradiation causes proliferating somatic cells in *C. elegans* larvae to arrest. This is visible by growth cessation, but not immediate death, of larvae. Like the embryonic lethality following irradiation of germ cells (Figure 3(a)), this arrest is an easy measure to test involvement of a protein in the UV response or NER (Figure 3(b)). NER deficient larvae are extremely sensitive to UV irradiation and arrest permanently at low doses of UV [63, 71]. Interestingly and in contrast to germ cells, TC-NER is the main repair pathway that is essential for survival and counteracts the growth arrest after UV irradiation [71]. GG-NER seems to be less important and only becomes significant for survival if TC-NER is deficient. Unlike in germ cells, this likely reflects the fact that only actively expressed genes need to be maintained in the mortal somatic cell lineages.

Permanent developmental arrest induced by UV irradiation does probably not depend on activation of checkpoint signaling, as knockdown of known UV-response checkpoint proteins does not alleviate arrest [63]. However, high UV doses induce permanent transcription block of a reporter gene and degradation of RNA polymerase II (*ama*-1) in *xpa-1* animals. Therefore, developmental growth arrest is likely due to transcription inhibition. In adult animals, in which somatic cells do not proliferate any more, UV irradiation causes the animals to become smaller, feed less, and live shorter, which is severely increased in NER deficient animals [92]. This reduction of growth could also be caused by lack of transcription. Additionally, it was suggested that UV might inhibit endoreduplication of epidermal syncytium nuclei, which normally increases cell size and drives adult growth in *C. elegans* [100].

Besides an effect on growth and development, DNA damage might also affect aging of adult *C. elegans*, as it does in mammals. *C. elegans* is a commonly used model organism to study aging, that lives on average for 2-3 weeks (under lab conditions), in which the insulin/IGF1 pathway regulates lifespan as it does in other organisms [101]. In mammals, there is also strong evidence that aging is in part caused by stochastic accumulation of damage to biomolecules such as DNA, caused by various environmental and metabolic agents [38, 41, 102]. Several human progeroid syndromes are caused by defective DNA repair mechanisms, in particular NER.

There is some evidence that suggests that longevity and repair are also linked in *C. elegans*. For instance, long-lived mutants are more resistant to oxidative stress and UV irradiation [86, 103–105]. Furthermore, increased UV resistance [105] and NER activity itself [86] were reported to be dependent on the insulin/IGF1 pathway. Still, a clear view on the relation between DNA repair and aging is blurred because of several seemingly contradictory reports. An early examination by Hartman and coworkers of four inbred strains with different life spans found no correlation with DNA repair competence [59]. Furthermore, of all *rad* mutants, only *rad-2* had a severely shortened lifespan [106]. Conflicting data exist on the life span of *xpa-1* mutants. We and others (unpublished data; [63, 92]) find that *xpa-1* mutants have a similar lifespan as wild type, but others have shown a shorter lifespan for these mutants [86, 87, 106]. These differences may be due to differences in experimental procedures, such as temperature which affects lifespan or the use of FUDR which is applied to prevent egg laying during life span assays. FUDR blocks DNA synthesis causing genomic stress which may, therefore, have an unanticipated effect on lifespan of DNA repair mutants. Also, wild-type life spans differ in each laboratory, which may lead to different conclusions when comparing wild-type and *xpa-1* lifespans. Furthermore, human patients and mouse models with XPA deficiency do not show progeroid features. Therefore, in addition to *xpa-1,* other NER deficient mutants should be studied to determine whether or not a relationship between NER, DNA damage and aging exists in *C. elegans*. Another unresolved issue is the fact that aging in mammals is a process that takes place in organisms that still have many proliferative tissues and stem cells with the capacity for cell renewal, whereas the somatic tissues of aging *C. elegans* do not proliferate.

Even if NER deficiency by itself might not be sufficient to shorten lifespan of *C. elegans*, UV-mediated induction of DNA damage severely shortens life span. A single or daily low UV dose is extremely limiting to lifespan in *xpa-1* animals but not in NER proficient animals. A higher dose (e.g., $\geq 50\,J/m^2$ UV-C) also limits wild-type life span but still to a much lesser extent [86, 92]. In addition to a reduced lifespan, UV exposed *xpa-1* mutants also exhibit other features of aging, such as damaged tissues and internal vacuoles. Thus, a negative correlation between NER deficiency and life span shortening in *C. elegans* might be explained by the lack of sufficient DNA damage accumulation during the short life time of *C. elegans* compared to mammals. If more damage is artificially induced, life span is severely shortened. Like UV-induced embryonic lethality and larval arrest, this UV-induced lifespan reduction is sometimes used as a measure to test involvement of a protein in the UV response or NER (Figure 3(c)).

Whole genome expression profiling of NER-deficient progeroid mice has suggested that in rapid aging tissue growth hormone (insulin/IGF1) signaling is downregulated whereas antistress response pathways are upregulated [38, 41, 102, 107]. As this response is similar to what happens in aging tissue, it was suggested that this reflects a compensatory survival response to counteract aging [38]. Whole genome profiling of *C. elegans* under standard laboratory conditions did not show major differences (>4-fold up- or downregulated genes) between wild type and *xpa-1* animals [92]. However, using a different methodology and less stringent criteria for up- and downregulation ($\geq$1.8-fold), many transcripts were found to be differentially regulated between wild-type and *xpa-1* mutants [87]. Gene Ontology enrichment analysis showed a bias of differentially regulated genes belonging to biological pathway clusters such as adult lifespan determination, ER unfolded protein response, regulation of carboxylic acid metabolism and phosphate transport. This enrichment is reminiscent of the suppression of the somatotroph axis and upregulation of stress response pathways in XPA-deficient mouse dermal fibroblast [108], as well as NER-deficient mice [41, 102]. This might mean that some transcriptomic

changes associated with NER deficiency and aging might be conserved between *C. elegans* and mammals. Following UV irradiation, several genes and biological networks potentially involved in a stress response were found to be differentially regulated in a similar manner in wild-type and *xpa-1* animals [92]. Importantly, NER genes as well as most other DNA repair genes are not transcriptionally induced after UV irradiation in *C. elegans*, which is similar to the lack of strong transcriptional regulation in mammals [109] but in contrast to bacteria [110] and yeast [111]. In summary, there is evidence that links NER deficiency and aging in *C. elegans,* but some results are still ambiguous. So far, all studies have made use of the *xpa-1* mutant which, although it is completely NER deficient, in mice is not strongly associated with accelerated aging. Therefore, studies utilizing other NER-deficient animals may be necessary to deduce whether *C. elegans* can be used as model for the damage accumulation theory of aging.

8. *C. elegans* as Model to Study the UV-Induced DNA Damage Response

In conclusion, studies on different aspects of the UV response confirm the important role of NER in *C. elegans*. Importantly, recent studies using *C. elegans* represent excellent examples of the different DNA damage responses in distinct cell types [64, 71, 72]. These experiments also show that survival and growth following DNA damage are not necessarily linked. This is for instance also evident from the fact that NER-deficient Dauer larvae survive UV irradiation, but are incapable of resuming normal development [63]. Thus, an organism's response to UV irradiation depends very much on its developmental status, its proliferative capacity, and its different tissues and cells.

Many aspects of NER and its role in the UV-DDR are still not well understood. In addition to studying the *in vivo* context of NER in *C. elegans,* this organism is also well suited to genetically identify new NER or UV-DDR regulatory pathways. Although it is not expected that novel core NER genes will be identified, genetic screening of *C. elegans* might prove useful to better understand the context of proteins and pathways in which NER plays a role and by which NER is regulated. In a first attempt to identify such proteins, we have recently identified several ATP-dependent chromatin remodeling factors that are essential for an efficient UV response [71]. Further characterization of novel genes and the cellular responses to UV in *C. elegans* will undoubtedly help to better understand the function of this important DNA repair pathway and etiologies of DNA damage-associated diseases.

Acknowledgments

The authors thank Jan Hoeijmakers for advise and critical discussion. Furthermore, they acknowledge financial support from the Association for International Cancer Research (Project no. 08-0084), the Netherlands Organization for Scientific Research (Project no. 863.08.022), the Netherlands Organization for health research and development (Project no. 40-00812-98-08031), and the EU funded Network of Excellence LifeSpan (FP6 036894).

References

[1] D. S. Boss, J. H. Beijnen, and J. H. Schellens, "Inducing synthetic lethality using PARP inhibitors," *Current Clinical Pharmacology*, vol. 5, no. 3, pp. 192–195, 2010.

[2] J. Essers, W. Vermeulen, and A. B. Houtsmuller, "DNA damage repair: anytime, anywhere?" *Current Opinion in Cell Biology*, vol. 18, no. 3, pp. 240–246, 2006.

[3] S. P. Jackson and J. Bartek, "The DNA-damage response in human biology and disease," *Nature*, vol. 461, no. 7267, pp. 1071–1078, 2009.

[4] G. Giglia-Mari, A. Zotter, and W. Vermeulen, "DNA damage response," *Cold Spring Harbor Perspectives in Biology*, vol. 3, Article ID a000745, 2011.

[5] B. B. Lemmens and M. Tijsterman, "DNA double-strand break repair in *Caenorhabditis elegans*," *Chromosoma*, vol. 120, no. 1, pp. 1–21, 2010.

[6] N. O'Neil and A. Rose, "DNA repair," *WormBook*, pp. 1–12, 2006.

[7] J. L. Youds, L. J. Barber, and S. J. Boulton, "*C. elegans*: a model of Fanconi anemia and ICL repair," *Mutation Research*, vol. 668, no. 1-2, pp. 103–116, 2009.

[8] Y. Matsumura and H. N. Ananthaswamy, "Toxic effects of ultraviolet radiation on the skin," *Toxicology and Applied Pharmacology*, vol. 195, no. 3, pp. 298–308, 2004.

[9] J. Cadet, E. Sage, and T. Douki, "Ultraviolet radiation-mediated damage to cellular DNA," *Mutation Research*, vol. 571, no. 1-2, pp. 3–17, 2005.

[10] J. R. Mitchell, J. H. Hoeijmakers, and L. J. Niedernhofer, "Divide and conquer: nucleotide excision repair battles cancer and ageing," *Current Opinion in Cell Biology*, vol. 15, no. 2, pp. 232–240, 2003.

[11] S. Prakash and L. Prakash, "Nucleotide excision repair in yeast," *Mutation Research*, vol. 451, no. 1-2, pp. 13–24, 2000.

[12] L. C. Gillet and O. D. Scharer, "Molecular mechanisms of mammalian global genome nucleotide excision repair," *Chemical Reviews*, vol. 106, no. 2, pp. 253–276, 2006.

[13] P. C. Hanawalt and G. Spivak, "Transcription-coupled DNA repair: two decades of progress and surprises," *Nature Reviews Molecular Cell Biology*, vol. 9, no. 12, pp. 958–970, 2008.

[14] J. Q. Svejstrup, "Contending with transcriptional arrest during RNAPII transcript elongation," *Trends in Biochemical Sciences*, vol. 32, no. 4, pp. 165–171, 2007.

[15] M. Fousteri, W. Vermeulen, A. A. van Zeeland, and L. H. Mullenders, "Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo," *Molecular Cell*, vol. 23, no. 4, pp. 471–482, 2006.

[16] K. A. Henning, L. Li, N. Iyer et al., "The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH," *Cell*, vol. 82, no. 4, pp. 555–564, 1995.

[17] C. Troelstra, A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma, and J. H. Hoeijmakers, "ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes," *Cell*, vol. 71, no. 6, pp. 939–953, 1992.

[18] A. J. van Gool, R. Verhage, S. M. Swagemakers et al., "RAD26, the functional S.cerevisiae homolog of the cockayne

syndrome B gene ERCC6," *The EMBO Journal*, vol. 13, no. 22, pp. 5361–5369, 1994.

[19] M. Araki, C. Masutani, M. Takemura et al., "Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair," *Journal of Biological Chemistry*, vol. 276, no. 22, pp. 18665–18672, 2001.

[20] K. Sugasawa, J. M. Ng, C. Masutani et al., "Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair," *Molecular Cell*, vol. 2, no. 2, pp. 223–232, 1998.

[21] M. Wakasugi, A. Kawashima, H. Morioka et al., "DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair," *Journal of Biological Chemistry*, vol. 277, no. 3, pp. 1637–1640, 2002.

[22] R. Groisman, J. Polanowska, I. Kuraoka et al., "The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage," *Cell*, vol. 113, no. 3, pp. 357–367, 2003.

[23] K. Sugasawa, Y. Okuda, M. Saijo et al., "UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex," *Cell*, vol. 121, no. 3, pp. 387–400, 2005.

[24] H. Wang, L. Zhai, J. Xu et al., "Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage," *Molecular Cell*, vol. 22, no. 3, pp. 383–394, 2006.

[25] J. H. Min and N. P. Pavletich, "Recognition of DNA damage by the Rad4 nucleotide excision repair protein," *Nature*, vol. 449, no. 7162, pp. 570–575, 2007.

[26] S. N. Guzder, P. Sung, L. Prakash, and S. Prakash, "Affinity of yeast nucleotide excision repair factor 2, consisting of the Rad4 and Rad23 proteins, for ultraviolet damaged DNA," *Journal of Biological Chemistry*, vol. 273, no. 47, pp. 31541–31546, 1998.

[27] T. G. Gillette, S. Yu, Z. Zhou, R. Waters, S. A. Johnston, and S. H. Reed, "Distinct functions of the ubiquitin-proteasome pathway influence nucleotide excision repair," *The EMBO Journal*, vol. 25, no. 11, pp. 2529–2538, 2006.

[28] M. Volker, M. J. Mone, P. Karmakar et al., "Sequential assembly of the nucleotide excision repair factors in vivo," *Molecular Cell*, vol. 8, no. 1, pp. 213–224, 2001.

[29] M. Yokoi, C. Masutani, T. Maekawa, K. Sugasawa, Y. Ohkuma, and F. Hanaoka, "The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA," *Journal of Biological Chemistry*, vol. 275, no. 13, pp. 9870–9875, 2000.

[30] K. Sugasawa, J. Akagi, R. Nishi, S. Iwai, and F. Hanaoka, "Two-step recognition of DNA damage for mammalian nucleotide excision repair: directional binding of the XPC complex and DNA strand scanning," *Molecular Cell*, vol. 36, no. 4, pp. 642–653, 2009.

[31] W. L. de Laat, E. Appeldoorn, K. Sugasawa, E. Weterings, N. G. Jaspers, and J. H. Hoeijmakers, "DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair," *Genes and Development*, vol. 12, no. 16, pp. 2598–2609, 1998.

[32] J. C. Huang, D. L. Svoboda, J. T. Reardon, and A. Sancar, "Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 8, pp. 3664–3668, 1992.

[33] A. O'Donovan, A. A. Davies, J. G. Moggs, S. C. West, and R. D. Wood, "XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair," *Nature*, vol. 371, no. 6496, pp. 432–435, 1994.

[34] A. M. Sijbers, W. L. de Laat, R. R. Ariza et al., "Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease," *Cell*, vol. 86, no. 5, pp. 811–822, 1996.

[35] R. M. Overmeer, A. M. Gourdin, A. Giglia-Mari et al., "Replication factor C recruits DNA polymerase delta to sites of nucleotide excision repair but is not required for PCNA recruitment," *Molecular and Cellular Biology*, vol. 30, no. 20, pp. 4828–4839, 2010.

[36] J. Moser, H. Kool, I. Giakzidis, K. Caldecott, L. H. Mullenders, and M. I. Fousteri, "Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase IIIα in a cell-cycle-specific manner," *Molecular Cell*, vol. 27, no. 2, pp. 311–323, 2007.

[37] T. Ogi, S. Limsirichaikul, R. M. Overmeer et al., "Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells," *Molecular Cell*, vol. 37, no. 5, pp. 714–727, 2010.

[38] J. H. Hoeijmakers, "DNA damage, aging, and cancer," *The New England Journal of Medicine*, vol. 361, no. 15, pp. 1475–1485, 2009.

[39] L. J. Niedernhofer, "Nucleotide excision repair deficient mouse models and neurological disease," *DNA Repair*, vol. 7, no. 7, pp. 1180–1189, 2008.

[40] N. G. Jaspers, A. Raams, M. C. Silengo et al., "First reported patient with human ERCC1 deficiency has cerebro-oculo-facio- skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure," *The American Journal of Human Genetics*, vol. 80, no. 3, pp. 457–466, 2007.

[41] L. J. Niedernhofer, G. A. Garinis, A. Raams et al., "A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis," *Nature*, vol. 444, no. 7122, pp. 1038–1043, 2006.

[42] J. J. Sekelsky, M. H. Brodsky, and K. C. Burtis, "DNA repair in *Drosophila*: insights from the *Drosophila* genome sequence," *Journal of Cell Biology*, vol. 150, no. 2, pp. F31–F36, 2000.

[43] J. P. Mueller and M. J. Smerdon, "Rad23 is required for transcription-coupled repair and efficient overall repair in *Saccharomyces cerevisiae*," *Molecular and Cellular Biology*, vol. 16, no. 5, pp. 2361–2368, 1996.

[44] R. A. Verhage, A. M. Zeeman, M. Lombaerts, P. van de Putte, and J. Brouwer, "Analysis of gene- and strand-specific repair in the moderately UV-sensitive *Saccharomyces cerevisiae rad23*mutant," *Mutation Research*, vol. 362, no. 2, pp. 155–165, 1996.

[45] H. de Waard, E. Sonneveld, J. de Wit et al., "Cell-type-specific consequences of nucleotide excision repair deficiencies: embryonic stem cells versus fibroblasts," *DNA Repair*, vol. 7, no. 10, pp. 1659–1669, 2008.

[46] T. Nouspikel, "DNA repair in mammalian cells: nucleotide excision repair: variations on versatility," *Cellular and Molecular Life Sciences*, vol. 66, no. 6, pp. 994–1009, 2009.

[47] A. P. Eker, C. Quayle, I. Chaves, and G. T. van der Horst, "DNA repair in mammalian cells: direct DNA damage reversal: elegant solutions for nasty problems," *Cellular and Molecular Life Sciences*, vol. 66, no. 6, pp. 968–980, 2009.

[48] M. Budzowska and R. Kanaar, "Mechanisms of dealing with DNA damage-induced replication problems," *Cell Biochemistry and Biophysics*, vol. 53, no. 1, pp. 17–31, 2009.

[49] S. J. Boulton, A. Gartner, J. Reboul et al., "Combined functional genomic maps of the *C. elegans* DNA damage response," *Science*, vol. 295, no. 5552, pp. 127–131, 2002.

[50] H. Morinaga, S. I. Yonekura, N. Nakamura, H. Sugiyama, S. Yonei, and Q. M. Zhang-Akiyama, "Purification and characterization of *Caenorhabditis elegans* NTH, a homolog of human endonuclease III: essential role of N-terminal region," *DNA Repair*, vol. 8, no. 7, pp. 844–851, 2009.

[51] N. Nakamura, H. Morinaga, M. Kikuchi et al., "Cloning and characterization of uracil-DNA glycosylase and the biological consequences of the loss of its function in the nematode *Caenorhabditis elegans*," *Mutagenesis*, vol. 23, no. 5, pp. 407–413, 2008.

[52] A. Shatilla and D. Ramotar, "Embryonic extracts derived from the nematode *Caenorhabditis elegans* remove uracil from DNA by the sequential action of uracil-DNA glycosylase and AP (apurinic/apyrimidinic) endonuclease," *Biochemical Journal*, vol. 365, no. 2, pp. 547–553, 2002.

[53] N. P. Degtyareva, P. Greenwell, E. R. Hofmann et al., "*Caenorhabditis elegans* DNA mismatch repair gene msh-2 is required for microsatellite stability and maintenance of genome integrity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2158–2163, 2002.

[54] M. Tijsterman, J. Pothof, and R. H. A. Plasterk, "Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient *Caenorhabditis elegans*," *Genetics*, vol. 161, no. 2, pp. 651–660, 2002.

[55] P. S. Hartman and R. K. Herman, "Radiation-sensitive mutants of *Caenorhabditis elegans*," *Genetics*, vol. 102, no. 2, pp. 159–178, 1982.

[56] P. S. Hartman, "UV irradiation of wild type and radiation-sensitive mutants of the nematode *Caenorhabditis elegans*: fertilities, survival, and parental effects," *Photochemistry and Photobiology*, vol. 39, no. 2, pp. 169–175, 1984.

[57] P. S. Hartman, "Epistatic interactions of radiation-sensitive (RAD) mutants of *Caenorhabditis elegans*," *Genetics*, vol. 109, no. 1, pp. 81–93, 1985.

[58] P. S. Hartman, J. Hevelone, V. Dwarakanath, and D. L. Mitchell, "Excision repair of UV radiation-induced DNA damage in *Caenorhabditis elegans*," *Genetics*, vol. 122, no. 2, pp. 379–385, 1989.

[59] P. S. Hartman, V. J. Simpson, T. Johnson, and D. Mitchell, "Radiation sensitivity and DNA repair in *Caenorhabditis elegans* strains with different mean life spans," *Mutation Research*, vol. 208, no. 2, pp. 77–82, 1988.

[60] C. A. Jones and P. S. Hartman, "Replication in UV-irradiated *Caenorhabditis elegans* embryos," *Photochemistry and Photobiology*, vol. 63, no. 2, pp. 187–192, 1996.

[61] P. S. Hartman, "Effects of age and liquid holding on the UV-radiation sensitivities of wild-type and mutant *Caenorhabditis elegans* dauer larvae," *Mutation Research*, vol. 132, no. 3-4, pp. 95–99, 1984.

[62] S. Ahmed, A. Alpi, M. O. Hengartner, and A. Gartner, "*C. elegans* RAD-5/CLK-2 defines a new DNA damage checkpoint protein," *Current Biology*, vol. 11, no. 24, pp. 1934–1944, 2001.

[63] J. W. Astin, N. J. O'Neil, and P. E. Kuwabara, "Nucleotide excision repair and the degradation of RNA pol II by the *Caenorhabditis elegans* XPA and Rsp5 orthologues, RAD-3 and WWP-1," *DNA Repair*, vol. 7, no. 2, pp. 267–280, 2008.

[64] A. H. Holway, S. H. Kim, A. La Volpe, and W. M. Michael, "Checkpoint silencing during the DNA damage response in *Caenorhabditis elegans* embryos," *Journal of Cell Biology*, vol. 172, no. 7, pp. 999–1008, 2006.

[65] J. N. Meyer, W. A. Boyd, G. A. Azzam, A. C. Haugen, J. H. Freedman, and B. van Houten, "Decline of nucleotide excision repair capacity in aging *Caenorhabditis elegans*," *Genome Biology*, vol. 8, no. 5, article R70, 2007.

[66] A. Gartner, S. Milstein, S. Ahmed, J. Hodgkin, and M. O. Hengartner, "A conserved checkpoint pathway mediates DNA damage—induced apoptosis and cell cycle arrest in *C. elegans*," *Molecular Cell*, vol. 5, no. 3, pp. 435–443, 2000.

[67] N. Ishii, N. Suzuki, P. S. Hartman, and K. Suzuki, "The effects of temperature on the longevity of a radiation-sensitive mutant rad-8 of the nematode *Caenorhabditis elegans*," *Journals of Gerontology*, vol. 49, no. 3, pp. B117–B120, 1994.

[68] M. H. Lee, B. Ahn, I. S. Choi, and H. S. Koo, "The gene expression and deficiency phenotypes of Cockayne syndrome B protein in *Caenorhabditis elegans*," *FEBS Letters*, vol. 522, no. 1-3, pp. 47–51, 2002.

[69] H. K. Park, D. Suh, M. Hyun, H. S. Koo, and B. Ahn, "A DNA repair gene of *Caenorhabditis elegans*: a homolog of human XPF," *DNA Repair*, vol. 3, no. 10, pp. 1375–1383, 2004.

[70] H. K. Park, J. S. Yook, H. S. Koo, I. S. Choi, and B. Ahn, "The *Caenorhabditis elegans* XPA homolog of human XPA," *Molecules and Cells*, vol. 14, no. 1, pp. 50–55, 2002.

[71] H. Lans, J. A. Marteijn, B. Schumacher, J. H. Hoeijmakers, G. Jansen, and W. Vermeulen, "Involvement of global genome repair, transcription coupled repair, and chromatin remodeling in UV DNA damage response changes during development," *PLoS Genetics*, vol. 6, Article ID e1000941, 2010.

[72] L. Stergiou, K. Doukoumetzidis, A. Sendoel, and M. O. Hengartner, "The nucleotide excision repair pathway is required for UV-C-induced apoptosis in *Caenorhabditis elegans*," *Cell Death and Differentiation*, vol. 14, no. 6, pp. 1129–1138, 2007.

[73] L. Stergiou, R. Eberhard, K. Doukoumetzidis, and M. O. Hengartner, "NER and HR pathways act sequentially to promote UV-C-induced germ cell apoptosis in *Caenorhabditis elegans*," *Cell Death and Differentiation*, vol. 18, pp. 897–906, 2011.

[74] C. I. Keller, J. Calkins, P. S. Hartman, and C. S. Rupert, "UV photobiology of the nematode *Caenorhabditis elegans*: action spectra, absence of photoreactivation and effects of caffeine," *Photochemistry and Photobiology*, vol. 46, no. 4, pp. 483–488, 1987.

[75] Y. Kim and E. T. Kipreos, "The *Caenorhabditis elegans* replication licensing factor CDT-1 is targeted for degradation by the CUL-4/DDB-1 complex," *Molecular and Cellular Biology*, vol. 27, no. 4, pp. 1394–1406, 2007.

[76] J. Hodgkin, H. R. Horvitz, and S. Brenner, "Nondisjunction mutants of the nematode *Caenorhabditis elegans*," *Genetics*, vol. 91, no. 1, pp. 67–94, 1979.

[77] D. B. Pontier and M. Tijsterman, "A Robust network of double-strand break repair pathways governs genome integrity during *C. elegans* development," *Current Biology*, vol. 19, no. 16, pp. 1384–1388, 2009.

[78] T. T. Saito, J. L. Youds, S. J. Boulton, and M. P. Colaiacovo, "*Caenorhabditis elegans* HIM-18/SLX-4 interacts with SLX-1 and XPF-1 and maintains genomic integrity in the germline by processing recombination intermediates," *PLoS Genetics*, vol. 5, no. 11, Article ID e1000735, 2009.

[79] J. L. Youds, N. J. O'Neil, and A. M. Rose, "Homologous recombination is required for genome stability in the absence of DOG-1 in *Caenorhabditis elegans*," *Genetics*, vol. 173, no. 2, pp. 697–708, 2006.

[80] R. S. Kamath, A. G. Fraser, Y. Dong et al., "Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi," *Nature*, vol. 421, no. 6920, pp. 231–237, 2003.

[81] B. Sönnichsen, L. B. Koski, A. Walsh et al., "Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*," *Nature*, vol. 434, no. 7032, pp. 462–469, 2005.

[82] T. Ohkumo, C. Masutani, T. Eki, and F. Hanaoka, "Deficiency of the *Caenorhabditis elegans* DNA polymerase η homologue increases sensitivity to UV radiation during germ-line development," *Cell Structure and Function*, vol. 31, no. 1, pp. 29–37, 2006.

[83] J. Pothof, G. van Haaften, K. Thijssen et al., "Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi," *Genes and Development*, vol. 17, no. 4, pp. 443–448, 2003.

[84] M. R. Wallenfang and G. Seydoux, "cdk-7 is required for mRNA transcription and cell cycle progression in *Caenorhabditis elegans*embryos," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 8, pp. 5527–5532, 2002.

[85] T. Eki, T. Ishihara, I. Katsura, and F. Hanaoka, "A genome-wide survey and systematic RNAi-based characterization of helicase-like genes in *Caenorhabditis elegans*," *DNA Research*, vol. 14, no. 4, pp. 183–199, 2007.

[86] M. Hyun, J. Lee, K. Lee, A. May, V. A. Bohr, and B. Ahn, "Longevity and resistance to stress correlate with DNA repair capacity in *Caenorhabditis elegans*," *Nucleic Acids Research*, vol. 36, no. 4, pp. 1380–1389, 2008.

[87] O. Fensgard, H. Kassahun, I. Bombik, T. Rognes, J. M. Lindvall, and H. Nilsen, "A two-tiered compensatory response to loss of DNA repair modulates aging and stress response pathways," *Aging*, vol. 2, no. 3, pp. 133–159, 2010.

[88] K. Kiontke and W. Sudhaus, "Ecology of Caenorhabditis species," *WormBook*, pp. 1–14, 2006.

[89] E. C. Friedberg, G. C. Walker, W. Siede et al., *DNA Repair and Mutagenesis*, ASM Press, Washington, DC, USA, 2006.

[90] S. Brenner, "The genetics of *Caenorhabditis elegans*," *Genetics*, vol. 77, no. 1, pp. 71–94, 1974.

[91] D. L. Mitchell, C. A. Haipek, and J. M. Clarkson, "(6-4)photoproducts are removed from the DNA of UV-irradiated mammalian cells more efficiently than cyclobutane pyrimidine dimers," *Mutation Research*, vol. 143, no. 3, pp. 109–112, 1985.

[92] W. A. Boyd, T. L. Crocker, A. M. Rodriguez et al., "Nucleotide excision repair genes are expressed at low levels and are not detectably inducible in *Caenorhabditis elegans* somatic tissues, but their function is required for normal adult life after UVC exposure," *Mutation Research*, vol. 683, no. 1-2, pp. 57–67, 2010.

[93] D. Greenstein, "Control of oocyte meiotic maturation and fertilization," *WormBook*, pp. 1–12, 2005.

[94] T. L. Gumienny, E. Lambie, E. Hartwieg, H. R. Horvitz, and M. O. Hengartner, "Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline," *Development*, vol. 126, no. 5, pp. 1011–1022, 1999.

[95] M. Giannattasio, C. Follonier, H. Tourriere et al., "Exo1 competes with repair synthesis, converts NER intermediates to long ssDNA gaps, and promotes checkpoint activation," *Molecular Cell*, vol. 40, no. 1, pp. 50–62, 2010.

[96] J. A. Marteijn, S. Bekker-Jensen, N. Mailand et al., "Nucleotide excision repair-induced H2A ubiquitination is dependent on MDC1 and RNF8 and reveals a universal DNA damage response," *Journal of Cell Biology*, vol. 186, no. 6, pp. 835–847, 2009.

[97] P. Hartman, J. Reddy, and B. A. Svendsen, "Does translesion synthesis explain the UV-radiation resistance of DNA synthesis in *C. elegans* embryos?" *Mutation Research*, vol. 255, no. 2, pp. 163–173, 1991.

[98] C. Masutani, R. Kusumoto, A. Yamada et al., "The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta," *Nature*, vol. 399, no. 6737, pp. 700–704, 1999.

[99] S. H. Kim and W. M. Michael, "Regulated proteolysis of DNA polymerase η during the DNA-damage response in *C. elegans*," *Molecular Cell*, vol. 32, no. 6, pp. 757–766, 2008.

[100] E. Lozano, A. G. Sáez, A. J. Flemming, A. Cunha, and A. M. Leroi, "Regulation of growth by ploidy in *Caenorhabditis elegans*," *Current Biology*, vol. 16, no. 5, pp. 493–498, 2006.

[101] C. J. Kenyon, "The genetics of ageing," *Nature*, vol. 464, no. 7288, pp. 504–512, 2010.

[102] I. van der Pluijm, G. A. Garinis, R. M. Brandt et al., "Impaired genome maintenance suppresses the growth hormone—insulin-like growth factor 1 axis in mice with Cockayne syndrome," *PLoS Biology*, vol. 5, no. 1, p. e2, 2007.

[103] P. L. Larsen, "Aging and resistance to oxidative damage in *Caenorhabditis elegans*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 19, pp. 8905–8909, 1993.

[104] J. R. Vanfleteren, "Oxidative stress and ageing in *Caenorhabditis elegans*," *Biochemical Journal*, vol. 292, no. 2, part 2, pp. 605–608, 1993.

[105] S. Murakami and T. E. Johnson, "A genetic pathway conferring life extension and resistance to UV stress in *Caenorhabditis elegans*," *Genetics*, vol. 143, no. 3, pp. 1207–1218, 1996.

[106] T. E. Johnson and P. S. Hartman, "Radiation effects on life span in *Caenorhabditis elegans*," *Journals of Gerontology*, vol. 43, no. 5, pp. B137–B141, 1988.

[107] B. Schumacher, J. H. Hoeijmakers, and G. A. Garinis, "Sealing the gap between nuclear DNA damage and longevity," *Molecular and Cellular Endocrinology*, vol. 112, p. 117, 2009.

[108] G. A. Garinis, L. M. Uittenboogaard, H. Stachelscheid et al., "Persistent transcription-blocking DNA lesions trigger somatic growth attenuation associated with longevity," *Nature Cell Biology*, vol. 11, no. 5, pp. 604–615, 2009.

[109] P. C. Hanawalt, J. M. Ford, and D. R. Lloyd, "Functional characterization of global genomic DNA repair and its implications for cancer," *Mutation Research*, vol. 544, no. 2-3, pp. 107–114, 2003.

[110] C. Janion, "Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli," *International Journal of Biological Sciences*, vol. 4, no. 6, pp. 338–344, 2008.

[111] Y. Fu, L. Pastushok, and W. Xiao, "DNA damage-induced gene expression in *Saccharomyces cerevisiae*," *FEMS Microbiology Reviews*, vol. 32, no. 6, pp. 908–926, 2008.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 213824, 14 pages
doi:10.4061/2011/213824

Research Article

5′CAG and 5′CTG Repeats Create Differential Impediment to the Progression of a Minimal Reconstituted T4 Replisome Depending on the Concentration of dNTPs

Emmanuelle Delagoutte[1] and Giuseppe Baldacci[2]

[1] *Muséum National d'Histoire Naturelle, Département "Régulations, Développement et Diversité Moléculaire", Laboratoire de Régulations et Dynamique des Génomes, USM 0503—INSERM U 565—UMR 7196, Case Postale no 26, 57 rue Cuvier, 75231 Paris cedex 05, France*
[2] *Institut Jacques Monod, UMR7592, "Pathologies de la réplication de l'ADN", CNRS and Université Paris-Diderot, 15 Rue Hélène Brion, 75205 Paris Cedex 13,, France*

Correspondence should be addressed to Emmanuelle Delagoutte, emmanuelle.delagoutte@mnhn.fr

Received 3 February 2011; Accepted 13 May 2011

Academic Editor: Giuseppina Giglia-Mari

Instability of repetitive sequences originates from strand misalignment during repair or replicative DNA synthesis. To investigate the activity of reconstituted T4 replisomes across trinucleotide repeats (TNRs) during leading strand DNA synthesis, we developed a method to build replication miniforks containing a TNR unit of defined sequence and length. Each minifork consists of three strands, primer, leading strand template, and lagging strand template with a 5′ single-stranded (ss) tail. Each strand is prepared independently, and the minifork is assembled by hybridization of the three strands. Using these miniforks and a minimal reconstituted T4 replisome, we show that during leading strand DNA synthesis, the dNTP concentration dictates which strand of the structure-forming 5′CAG/5′CTG repeat creates the strongest impediment to the minimal replication complex. We discuss this result in the light of the known fluctuation of dNTP concentration during the cell cycle and cell growth and the known concentration balance among individual dNTPs.

1. Introduction

Repetitive sequences, such as trinucleotide repeats (TNRs), are spread all over the genome and can be found in intergenic regions, 5′ regulatory regions, promoters, introns, or exons. Such sequences are particularly prone to mutate, and their rate of mutation can be several orders of magnitude higher than that of bulk DNA [1]. A change in the repeat number of a given repetitive sequence can influence gene expression [2–4], allowing morphological evolution [5] or generation of diverse social behaviors [6]. On the other hand, a dozen genetic diseases (e.g., myotonic dystrophy, Huntington's disease, and a variety of ataxias) are caused by expansion of a TNR sequence located at a specific locus of the genome [7–11]. Similarly, a variety of cancers are due to frameshift mutations in the repetitive sequence of a given gene [12, 13]. The molecular mechanism underlying repetitive sequence instability has not been completely unraveled, and current models are based on strand slippage during replicative or repair DNA synthesis. Strand slippage is made possible by the repetitive nature of the sequence and leads to the formation of a hairpin on the template or newly synthesized strand. A recent study has provided direct evidence for hairpin formation during TNR replication *in vivo* [14] although the precise time at which hairpin forms is still unknown. Many of the *in vitro* studies published so far in the field of TNR instability relied on primer extension assays combining TNRs with different sequences and lengths and various DNA polymerases. If primer extension assays are good models of gap repair DNA synthesis, they are not fully adapted to study replicative DNA synthesis. The use of replication miniforks containing a defined TNR sequence on which a reconstituted

replisome of various complexities can assemble should be very helpful in dissecting the mechanism involved in TNR instability during DNA replication. For instance, the activity of each component of the replisome can be easily examined and the experimental conditions changed without difficulty. The T4 replication machinery is an ideal system, since it is the simplest replication system that uses essentially the same set of regulatory proteins also used by higher organisms (for reviews, see [15–20]). It requires only seven proteins to initiate and catalyze coordinated *in vitro* leading and lagging strand DNA synthesis with a speed, processivity, and fidelity similar to those measured *in vivo*. Among the T4 replisomal proteins are (1) the DNA polymerase, the gene product (gp) 43, with a 5′ → 3′ DNA synthesis activity and a 3′ → 5′ exonuclease or proofreading activity that removes misincorporated dNMPs, (2) the helicase, gp41, with a 5′ → 3′ double-stranded (ds) DNA strand separation activity that unwinds the parental duplex by sterically excluding the ss leading strand template and encircling the ss lagging strand template, (3) the processivity factor, gp45, a homotrimeric ring-shaped and noncatalytic protein that confers processivity to the DNA polymerase, (4) the clamp loader, gp44/62, a hetero-oligomeric complex that loads the processivity factor at a primer-template (p-t) junction, (5) the single-stranded (ss) DNA binding protein (SSB), gp32, that binds and protects the naked ss DNA exposed by the helicase, and finally (6) the primase, gp61, that synthesizes small RNA primers complementary to the ss lagging strand template that the lagging strand DNA polymerase extends. An eighth T4-encoded protein called the helicase loader protein, gp59, is required to load the helicase on ss DNA covered by SSB.

In this paper, we present a method to build replication miniforks containing a TNR of defined sequence and length. Each minifork consists of three strands, namely, primer, leading, and lagging strand template. The primer anneals to the 3′ end of the leading strand template to create a specific binding site, a p-t junction, for the DNA polymerase. The lagging strand template is complementary to the leading strand template except in its 5′ extremity. The ss tail at the 5′ end of the lagging strand template constitutes the assembly site of the replicative helicase. The method relies on the preparation of each strand of the minifork separately followed by the hybridization of the three strands of the minifork. The DNA synthesis activity of simple reconstituted bacteriophage T4 replisomes across specific ds TNR sequences was characterized. We show that a minimal replisome constituted by the T4 helicase (gp41) and the T4 DNA polymerase (gp43) can replicate the leading strand template of random, structure-forming or nonstructure-forming TNR sequences. The T4 helicase loader (gp59) increases the efficiency of strand displacement DNA synthesis of the minimal homologous couple and may reduce slippage of the helicase that is, hydrolysis of ATP nonassociated with ds DNA unwinding. In contrast, a heterologous couple constituted by the T4 helicase and Klenow fragment undergoes uncoordinated leading strand DNA synthesis, even in the presence of the T4 helicase loader. Coupled DNA synthesis at various concentrations of dNTPs was performed and surprisingly revealed that the strand of the structure-forming 5′CAG/5′CTG TNR that hinders the progression of the reconstituted T4 replication trio (gp41-gp43-gp59) depends on the concentration of dNTPs. For instance, at high [dNTPs], the greatest impediment to progression of the T4 replication trio is measured when the 5′CAG sequence is on the leading strand template. In contrast, at low [dNTPs], the 5′CTG sequence on the leading strand template creates a greater impediment to T4 replication trio progression than the 5′CAG sequence due to a very low efficiency of incorporation of dAMP across TMP of the 5′CTG repeat. The functional consequences of these results are discussed.

2. Materials and Methods

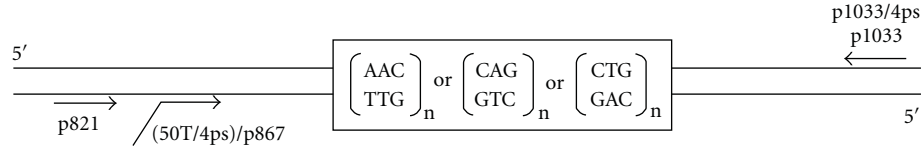
2.1. Enzymes. Herculase-enhanced DNA polymerase was from Stratagene; Phusion high-fidelity DNA polymerase was from Finnzymes; T7 exonuclease, T4 polynucleotide kinase (PNK), and Klenow fragment were from New England Biolabs (NEB). T4 helicase (gp41), T4 DNA polymerase (gp43), and T4 helicase loader (gp59) were prepared as described [21–25]. The proteins purified in our laboratory were estimated to be at least 90% pure by Coomassie Blue staining. Their concentration was measured by UV spectroscopy using an extinction coefficient of 7.6×10^4, 1.3×10^5, $3.8 \times 10^4\ M^{-1}cm^{-1}$ for monomeric gp41, monomeric gp43 and monomeric gp59, respectively.

2.2. DNAs. The three plasmids derived from pcDNA3 (Invitrogen) and containing a TNR of defined length and sequence are listed and described in Table 1. The plasmid p-Empty also derives from pcDNA3 but contains no repetitive sequence. Plasmid constructions are named based on the sequence of the leading template strand. Oligonucleotides were from Eurogentec and their sequence is listed in Table 1 as is also their annealing position with respect to the repetitive sequence. The oligonucleotides (50T/4ps)/p867 and 1033/4ps carry four phosphorothioate linkages at their 5′ extremity that make them resistant to the 5′ → 3′ T7 exonuclease. All oligonucleotides were gel-purified before use. Their concentrations were determined by UV spectroscopy using the extinction coefficients provided by the manufacturer.

2.3. PCR Conditions. Polymerase chain reactions (PCRs) were performed in the buffer of the DNA polymerase provided by the manufacturer. The concentrations of plasmid, primers, dNTPs, and enzyme were as recommended by the manufacturer, except for the oligonucleotide that contains the phosphorothioate linkages whose concentration is half the concentration recommended by the manufacturer. The cycling conditions when using the Phusion high-fidelity DNA polymerase were as follows: initial denaturation (3 min at 98°C); 25 cycles of the three following steps (10 sec at 98°C, 10 sec at 63°C, 15 sec at 72°C); final extension (10 min at 72°C). The cycling conditions when using the Herculase-enhanced DNA polymerase were as follows: initial denaturation (3 min at 98°C), and 25 cycles of the three following steps (40 sec at 98°C, 30 sec at 60°C, and 30 sec at

Table 1: Description of plasmids and oligonucleotides.

p-Empty	Plasmid that does not contain any repeat
p-5′GTT16	Plasmid that contains 16 repeats of 5′-AAC/5′-TTG
p-5′CTG17	Plasmid that contains 17 repeats of 5′-CAG/5′-GTC
p-5′CAG23	Plasmid that contains 23 repeats of 5′-CTG/5′-GAC
p821	5′-CTGGCTAACTAGAGAACCCACTGCTTACTGGC-3′
p1033	5′-GATCAGCGAGCTCTAGCATTTAGGTGACAC-3′
p1033/4ps	5′-G◊A◊T◊C◊AGCGAGCTCTAGCATTTAGGTGACAC-3′
(50T/4ps)/p867	5′-T◊T◊T◊T◊$(T)_{46}$CGACTCACTATAGGGAGACCCAAGCTAATT-3′
	◊indicates the position of the four phosphorothioate linkages within the oligonucleotide



The upper part of the table describes the plasmids. The middle part of the table refers to the oligonucleotides and their sequences. The lower part of the table shows the hybridization positions of the oligonucleotides.

72°C), and final extension (10 min at 72°C). After PCR, the Nucleospin extract kit (Macherey-Nagel) was used to remove unused primers as recommended by the manufacturer. The PCR products were analyzed by electrophoresis on a 2% agarose gel in 1X TBE.

2.4. T7 Exonuclease Digestion. T7 exonuclease digestion was performed in 1X NEB buffer 4 (50 mM KOAc, 20 mM Tris-OAc, 10 mM $MgOAc_2$ and 1 mM DTT pH 7.9 at 25°C) at 25°C. Prior to large scale digestion, the enzyme concentration and the duration of the digestion were adjusted by performing a small-scale digestion. The large-scale T7 exonuclease digestion was stopped by adding EDTA (final concentration of 80 mM) and proteinase K (final concentration of 1.5 mg/mL), and incubated for 10 min at 37°C and 10 min at 65°C. T7 digestion was checked by electrophoresis on a 2% agarose gel in 1X TBE.

2.5. Radiolabelling of p821. Radiolabelling of the oligonucleotide p821 at its 5′ extremity was performed by incubating 1 μM oligonucleotide with 1 μM $\{\gamma^{32}P\}$-ATP, 1 μM ATP and PNK (0.2 u/mL) in 1X PNK buffer (70 mM Tris-HCl, 10 mM $MgCl_2$ and 5 mM DTT pH 7.6 at 25°C) for 45 min at 37°C. Nonincorporated $\{\gamma^{32}P\}$-ATP and ATP were removed using a Biospin 6 column (Biorad) equilibrated in TE buffer.

2.6. Preparation of the Minifork. Miniforks were prepared by mixing the radiolabelled p821 primer with the ss leading and lagging strand templates in a buffer containing 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl. Strand hybridization was performed by heating (5 min at 95°C) and slow cooling. Hybridization was checked by electrophoresis on a native polyacrylamide gel. The mass ratio of acrylamide to bisacrylamide was 29 : 1 and native gels were in 1X TBE.

2.7. DNA Synthesis Assay. Except for Klenow fragment, protein concentrations are given in monomeric units. DNA synthesis was performed in the buffer used to prepare the miniforks supplemented with 250 μM each dNTP (unless indicated otherwise), 2.5 mM ATP, and 2.5 mM DTT. The DNA, gp43, gp41 and gp59 concentrations were 10 nM, 30 nM, 200 nM, and 200 nM, respectively. Klenow fragment was either at 1 mU/μL or 10 mU/μL as indicated in the figure. DNA and DNA polymerase (gp43 or Klenow fragment) were preincubated for 3 min at 37°C before adding gp41 premixed or not with gp59. At the indicated times, the reaction was quenched by adding EDTA to 50 mM. Proteinase K and SDS were added to a final concentration of 3 mg/mL and 0.05%, respectively, and proteolysis performed for 20 min at 37°C. One volume of denaturing blue (100% formamide, 0.01% bromophenol blue, and 0.01% xylene cyanol) was added to the samples that were then heated for 5 min at 95°C and loaded onto a 10% acrylamide sequencing gel (mass ratio of acrylamide to bisacrylamide = 19 : 1) in 1X TBE. After electrophoresis, the gel was dried and exposed on a phosphorimager screen. After at least 10 hours of exposure, the screen was scanned with a Storm 820 (GE Healthcare). The samples in the gel were quantified using ImageQuant version 5.1 or NT.

2.8. DNA Sequencing of the Leading Strand Template. The DNA substrates used for sequencing the ss leading strand templates were p-t junctions prepared in the 1X Sequenase Version 2.0 reaction buffer (40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$ and 50 mM NaCl) and consisted of the radiolabelled p821 annealed to the ss leading strand template. The p-t junctions were treated for sequencing as recommended in the Sequenase Version 2.0 T7 DNA polymerase kit (Usb).

2.9. ATPase Assay. Assays were performed in the buffer used to prepare the miniforks supplemented with 250 μM of each dNTP, 2.5 mM ATP, 0.1 μM $\{\gamma^{32}P\}$-ATP and 2.5 mM DTT. The DNA, gp43, gp41 and gp59 concentrations were 10 nM, 30 nM, 200 nM and 200 nM, respectively (concentrations

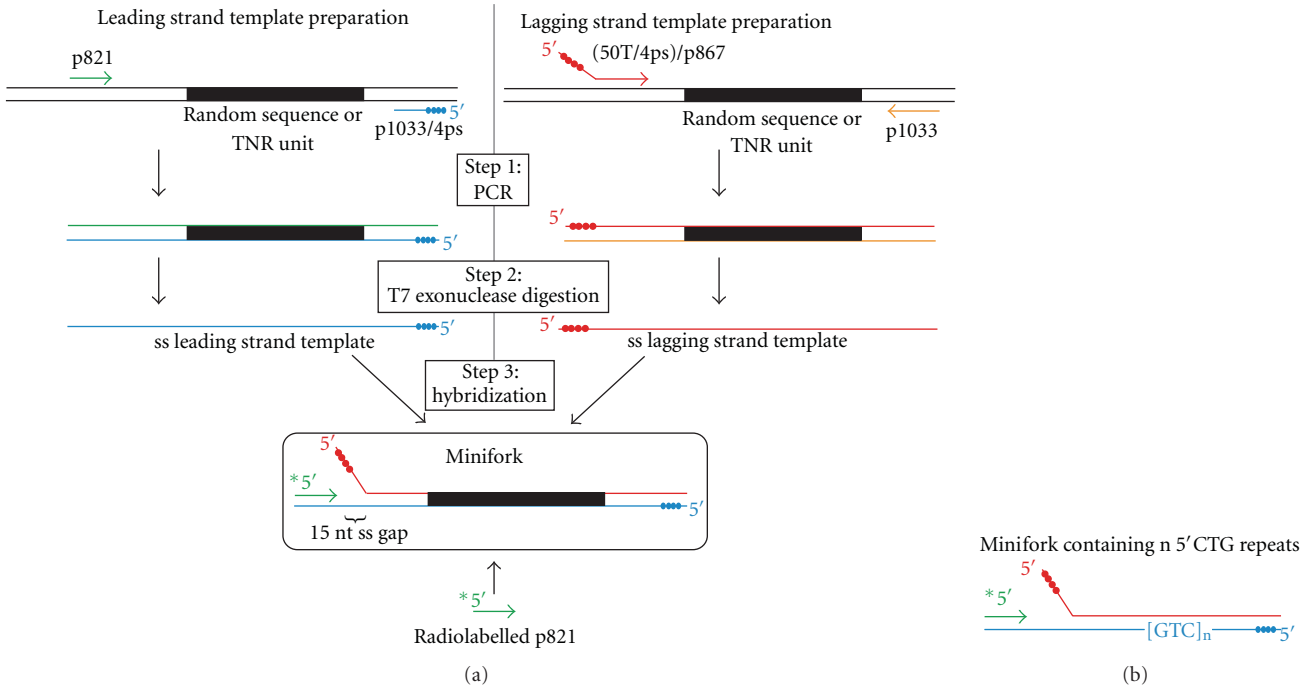


Figure 1: Strategy of the preparation of model replication miniforks. (a) The left and right sides of the figure correspond to the strategy used to build the ss leading and lagging strand templates of the minifork, respectively. The ss leading and lagging strand templates are combined with the radiolabelled p821 primer to assemble the minifork by strand hybridization. Replication miniforks are prepared in three consecutive steps. The first step (Step 1: PCR) consists of a PCR using plasmids containing a random or a TNR sequence (shown as a black rectangle) and oligonucleotides that flank the random sequence or the TNR unit. For each PCR, one of the oligonucleotides (p1033/4ps (colored in blue) for the preparation of the ss leading strand template, and (50T/4ps)/p867 (colored in red) for the preparation of the ss lagging strand template) carries 4 phosphorothioate linkages (represented as filled blue and red spheres for p1033/4ps and (50T/4ps)/p867, resp.) at its 5′ end. After PCR, the ds PCR products are digested by the T7 exonuclease that specifically degrades the DNA strand (colored in green or orange for the preparation of the ss leading or lagging strand template, resp.) that does not contain the phosphorothioate linkages (Step 2: T7 exonuclease digestion). The minifork (shown in a rounded rectangle) is assembled by hybridization of the ss leading and lagging strand templates and the radiolabelled p821 primer (in green) (Step 3: Hybridization). A gap of 15 nts exists between the 3′ end of the p821 primer and the base of the ss tail of the lagging strand template to facilitate the assembly of the DNA polymerase at the p-t junction. (b) A minifork containing n repeats of 5′CTG is shown. The repeats are located on the leading strand template.

expressed in monomeric units). When present, gp43 was preincubated with the minifork for 3 min at 37°C before adding gp41 premixed or not with gp59. At the indicated times the reaction was quenched by spotting an aliquot of the reaction on a PEI cellulose thin layer chromatography (TLC) plate. ATP and inorganic phosphate were separated by running the TLC plate in 0.35 M potassium phosphate buffer (pH 3). The TLC plate was next air-dried and exposed on a phosphorimager screen. After at least 10 hours of exposure, the screen was scanned with a Storm 820 (GE Healthcare). Radioactive ATP and inorganic radioactive phosphate were quantified using ImageQuant version 5.1 or NT.

3. Results

3.1. Strategy. The strategy used to build replication miniforks of defined sequences is shown in Figure 1(a) and can be divided into three steps. The ss leading and lagging strand templates are prepared separately by a two-step procedure. In a first step, a PCR is performed with a plasmid containing a random or a TNR sequence (Table 1) and two oligonucleotides, one of them carrying a track of four phosphorothioate linkages at its 5′ extremity to make it resistant to the 5′ → 3′ T7 exonuclease. The oligonucleotide couples for the leading and lagging strand templates are (p821, p1033/4ps) and ((50T/4ps)/p867, p1033), respectively (see Table 1 for oligonucleotide sequences and annealing positions). In a second step, the PCR fragments are treated with T7 exonuclease to specifically degrade the strand that is devoid of phosphorothioate linkage. Miniforks are assembled in a third step by annealing the radiolabelled p821 primer, the ss leading and the lagging strand templates by a heating and slow cooling procedure. Each minifork (shown in a rounded rectangle in Figure 1(a)) carries a p-t junction on which the DNA polymerase can bind to initiate DNA synthesis and a 5′ ss tail on which the helicase with or without the assistance of the helicase loader can assemble. The 15 nucleotide (nt) gap of ss DNA that exists between the 3′ end of the p821 primer and the base of the ss tail of the lagging strand template (Figure 1(a)) makes it possible for the DNA polymerase to assemble and initiate DNA synthesis by filling the 15 nt-long gap of ss DNA before starting

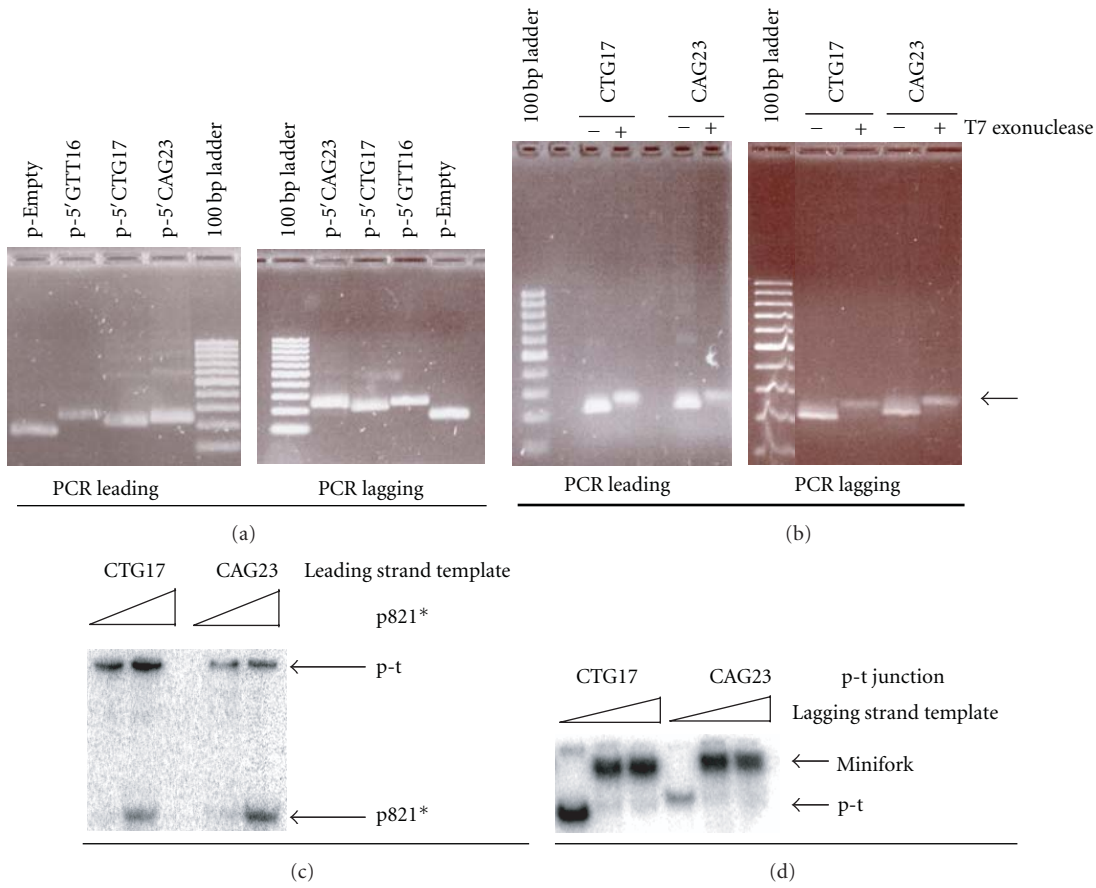


Figure 2: Preparation of leading and lagging strand templates, p-t junctions and miniforks. The DNA fragments in the 100 base pair (bp) ladder are 100, 200, 300, 400, 500/517, 600, 700, 800, 900, and 1000 pbs. (a) The agarose gels stained by ethidium bromide (EtBr) show the products of the PCR obtained with the four plasmids (p-Empty, p-5′GTT16, p-5′CTG17, and p-5′CAG23) and the oligonucleotide couple specific of the leading ("PCR leading"; left) or the lagging ("PCR lagging"; right) strand. The name of the plasmids used for the PCR is indicated at the top of the gels. (b) The agarose gels stained by EtBr show the products of the T7 exonuclease digestion. The PCR products obtained with p-5′CTG17 (CTG17) and p-5′CAG23 (CAG23) and the oligonucleotide couple specific of the leading ("PCR leading"; left) or the lagging ("PCR lagging"; right) strand were treated (+) or not (−) by T7 exonuclease. After treatment with T7 exonuclease, the appearance of a DNA band with a slower electrophoretic migration and a weaker intensity (indicated by a backward arrow) than the ds DNA is indicative of ss DNA production. (c) The ss leading strand templates containing either 17 repeats of CTG (CTG17) or 23 repeats of (CAG23) are mixed with increasing amounts of radiolabelled p821 (p821*) to generate the p-t junctions. Species are resolved on a native gel. Free p821 migrates faster than the p-t junctions. (d) The p-t junctions containing 17 repeats of CTG (CTG17) or 23 repeats of (CAG23) on their leading strand are mixed with increasing amounts of ss lagging strand template to assemble the miniforks. Species are resolved on a native gel. The miniforks migrate more slowly than the p-t junctions.

coupled DNA synthesis with the helicase. In what follows, the name of the TNR associated with a minifork refers to the TNR sequence of the leading strand template. For instance, the minifork containing a 5′CTG repeat has the 5′CTG repeat unit located on its leading strand template (Figure 1(b)).

3.2. Preparation of Leading and Lagging Templates of the Miniforks. Two different DNA polymerases were tested, the Herculase-enhanced DNA polymerase and the Phusion high-fidelity DNA polymerase, and both gave suitable results. To use all the oligonucleotide that carried the phosphorothioate linkages, its concentration was half the concentration recommended by the manufacturer. Each plasmid (p-Empty, p-5′GTT16, p-5′CTG17, and p-5′CAG23; Table 1) was used with the oligonucleotide couple specific of the leading or lagging strand template. A PCR fragment of expected size was synthesized in each case (Figure 2(a)). The PCR fragments were next treated with T7 exonuclease to generate ss DNA. As shown in Figure 2(b) (data shown for the production of the ss leading templates containing 17 repeats of CTG and 23 repeats of CAG (left panel) and the ss lagging strand templates containing 17 repeats of CAG and 23 repeats of CTG (right panel)), a DNA band with a slower mobility and a weaker intensity than the untreated ds PCR fragment appeared after T7 exonuclease treatment, indicative of the production of ss DNA.

3.3. Assembly of the Miniforks. All miniforks were rendered radioactive by the use of the radiolabelled p821 oligonucleotide. A two-step procedure was followed to assemble

the miniforks. First, radiolabelled p821 was annealed to the ss leading strand template to generate a p-t junction. Hybridization of the two DNAs was checked by electrophoresis on a native gel, because the free p821 primer and the p-t junction do not have the same electrophoretic mobility. Free p821 indeed migrated faster than the p-t junctions (Figure 2(c)). Second, the ss lagging strand template was annealed to the p-t junction, and strand hybridization similarly checked by electrophoresis on a native gel (Figure 2(d)).

3.4. Coupled DNA Synthesis across the ds Random Sequence of the Minifork. To test the quality of the miniforks, we first characterized the activity of the T4 DNA polymerase (gp43) alone or assisted by the T4 helicase (gp41) and T4 helicase loader (gp59) on a minifork carrying a random sequence. Gp43 was incubated for 3 min with the minifork before adding gp41 premixed or not with gp59 to initiate strand displacement DNA synthesis. As expected, during the first 3 min, gp43 filled the ss gap of the minifork up to the base of the ss lagging tail (Figure 3(A), lane 2). Gp43 was unable to efficiently synthesize through the DNA duplex due to a very weak intrinsic strand displacement activity, and a DNA intermediate corresponding to the p821 primer extended by 15 nts (labelled (b) in Figure 3(A)) accumulated. Addition of gp41 allowed coupled DNA synthesis across the duplex part of the minifork to take place (Figure 3(A), lanes 3–8) as pointed out by the accumulation of the full-length product (labelled (c) in Figure 3(A)) over time. Coupled leading strand DNA synthesis was highly stimulated by gp59 as a significantly higher amount of full length product accumulated over time when gp59 was part of the reconstituted replisome (Figure 3(A), lanes 9–15). In addition, in the presence of gp43, gp41, and gp59, three intermediate DNA synthesis products (indicated by a backward arrow in Figure 3(A)) near the end of the parental DNA duplex transiently accumulated to a significant extent; they might correspond to pause sites for the minimal reconstituted replisome from where DNA synthesis successfully resumed.

Similarly, the DNA synthesis pattern was examined in the presence of Klenow fragment. Klenow fragment has weak strand displacement DNA synthesis activity that allowed it to synthesize through the duplex part of the minifork (Figure 3(B), lanes 2 and 5). However, contrary to the gp43-gp41 replication couple, DNA synthesis performed by Klenow fragment was not stimulated by gp41 (Figure 3(B), compare lanes 2, 3, 5, and 6). Under both conditions (with or without gp41), DNA synthesis was highly distributive as indicated by the numerous pause sites that were clearly visible along the template strand. The addition of gp59 to the Klenow fragment-gp41 replication couple stimulated DNA synthesis (Figure 3(B), compare lanes 3, 4, 6, and 7), but DNA synthesis remained distributive, suggesting uncoordinated DNA synthesis with this heterologous replication system.

3.5. At High Concentrations of dNTPs, a 5′CAG Repeat Creates a Greater Impediment to the T4 gp41-gp43-gp59 Replication Trio than a 5′CTG Repeat. Miniforks containing 16 repeats of 5′GTT, 17 repeats of 5′CTG, or 23 repeats of 5′CAG on their leading strand template were built and the activity of minimal reconstituted T4 replisomes was tested on these miniforks. Similarly to the minifork containing a random sequence, the gp41-gp43 replication couple synthesized DNA quite efficiently across a nonstructure-forming (e.g., 5′GTT) or structure-forming TNRs (e.g., 5′CAG and 5′CTG) (Figure 4(A), result shown only for the minifork containing 17 repeats of 5′CTG, lanes 3 and 4). Addition of gp59 stimulated leading strand DNA synthesis fivefold (Figure 4(A), compare lanes 3, 6, 4, and 7). A similar stimulation of DNA synthesis by gp59, was measured with miniforks containing 16 repeats of 5′GTT or 23 repeats of 5′CAG (data not shown). The effect of gp59 on the ATPase activity of gp41 during coupled leading strand DNA synthesis was also evaluated. The gp41 ATPase activity during coupled leading strand DNA synthesis was measured in the presence or absence of gp59 and was compared to the activity measured in the absence of gp43 (Figure 5, result shown only for the minifork containing 17 repeats of 5′CTG). As expected [24], in the presence of the minifork as nucleic acid cofactor, gp59 stimulated the gp41 ATPase activity in the absence of gp43 (Figure 5, compare black and blue curves). This stimulation level is higher than reported in the absence of ss DNA ($\leq$2, [24]), indicating that under our experimental conditions, gp59 stimulates both the intrinsic and the ss DNA-dependent ATPase activity of gp41. Gp43 by itself was also able to stimulate the gp41 ATPase activity (Figure 5, compare black and red curves), but to a slightly lower extent than gp59 (Figure 5, compare red and blue curves). Nevertheless, the addition of gp59 to the gp41-gp43 replication couple led to a gp41 ATPase activity comparable to the one measured for the gp41-gp59 couple (Figure 5, compare green and blue curves). A similar trend of stimulation of the gp41 ATPase activity by gp59 was measured with miniforks containing 16 repeats of 5′GTT or 23 repeats of 5′CAG (data not shown). The fact that gp59 stimulates the DNA synthesis activity of the gp41-gp43 replication couple to a higher extent than it does the ATPase activity (fivefold (Figure 4(A), compare lanes 3, 6, 4, and 7) *versus* twofold (Figure 5, compare red and green curves)) possibly reflects a reduced amount of slippage of gp41 in the presence of gp59 during coupled strand displacement DNA synthesis.

DNA synthesis activity of the reconstituted gp41-gp43-gp59 replication trio across the 5′CAG and 5′CTG repeat unit was quantified. The rates of the minimal T4 replisome prior to, and across the TNR unit were estimated using the method previously described [26]. Briefly, the speed of the minimal reconstituted replisome before the TNR was determined by estimating the loss of signal intensity of primers extended by 14 to 16 nts (indicated by the bracket labelled "prior to TNR" in Figure 4(A)). The intensity of primers of these lengths decreased with time, since they were intermediate DNA synthesis products of larger extension products. To assess the effects of passage through the TNR region, the amount of elongated primers after passage of the minimal reconstituted replisome was measured. The amount of elongation products that contained the full length TNR unit (indicated by the bracket labelled "past TNR") increased over

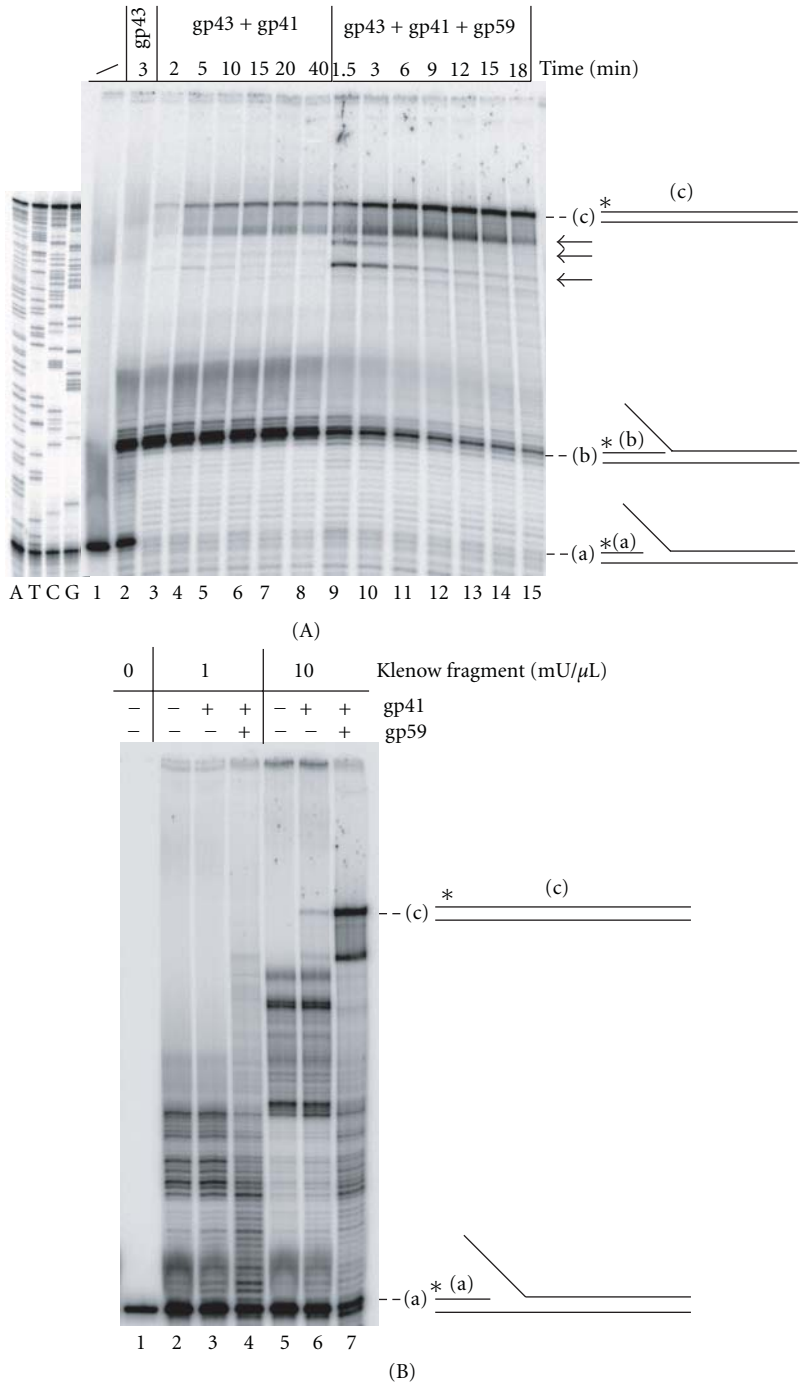


Figure 3: DNA synthesis activity of minimal reconstituted T4 replisomes across a minifork containing a random sequence. The minifork was prepared with the plasmid p-Empty and thus contains a random sequence to be replicated. gp43, gp41, and gp59 are the DNA polymerase, the helicase and the helicase loader of bacteriophage T4, respectively. (A) Gp43 was incubated with the minifork, alone (lane 2), with gp41 (lanes 3–8) or with gp41 and gp59 (lanes 9–15). The reaction was quenched at various times and the samples were loaded on a denaturing sequencing gel. (a) corresponds to the radiolabelled p821 primer. (b) corresponds to radiolabelled p821 extended by 15 nts up to the base of the 5′ ss tail of the lagging strand template. (c) corresponds to the radiolabelled p821 extended up to the end of the leading strand template after strand displacement DNA synthesis. The (a), (b), and (c) DNAs are also shown in the context of the minifork on the right side of the figure. Major transient intermediate products are indicated by backward arrows. The four sequencing reactions (A, T, C, and G) of the leading strand template of the minifork are shown on the left side of the figure. (B) The minifork was incubated with Klenow fragment alone (lanes 2 and 5), together with gp41 (lanes 3 and 6) or with gp41 and gp59 (lanes 4 and 7) for 20 minutes. The reaction products were resolved on a denaturing sequencing gel. Two amounts of Klenow fragment were tested (1 mU/μL, lanes 2–4; 10 mU/μL, lanes 5–7). (a) and (c) are as in 3A.

time, since these products accumulated as coupled leading strand DNA synthesis proceeded. The results indicated that before reaching the TNR unit, the reconstituted gp41-gp43-gp59 replication trio assembled on the minifork containing a 5′CAG unit synthesized DNA at a speed very similar to that of the reconstituted gp41-gp43-gp59 replication trio assembled on the minifork containing a 5′CTG unit (Figures 4(B) and 4(D)). However, replicating through a 5′CAG repeat created a greater impediment to the minimal T4 replisome than replicating through a 5′CTG sequence. The amount of synthesized DNA containing the full-length TNR unit was indeed lower when the gp41-gp43-gp59 replication complex replicated through a 5′CAG template than when it replicated through the 5′CTG repeat unit (Figures 4(C) and 4(D)). This result is in a perfect agreement with the relative degree of impediment created by 5′CAG and 5′CTG repeats when these TNR sequences are presented to gp43 in a ss context [26].

3.6. At Low Concentrations of dNTPs, a 5′CAG Repeat Creates a Weaker Impediment to the Minimal Reconstituted T4 Replisome than a 5′CTG Repeat. At high dNTP concentrations, the reconstituted gp41-gp43-gp59 replication trio copied quite efficiently a leading strand template containing random, structure-forming or nonstructure-forming TNR sequences, and full-length product accumulated over time. However, such experimental conditions did not provide the opportunity to characterize the activity of the minimal replisome across the TNR unit itself. Therefore, the speed of the gp41-gp43-gp59 replication trio was reduced by lowering the concentrations of the dNTPs and the DNA synthesis pattern was examined under these conditions. As expected, reducing the concentrations of dNTPs decreased the size of the DNA products that were synthesized by the minimal reconstituted T4 replisome during leading strand DNA synthesis (Figure 6(a)). The DNA synthesis profile prior to, across, and beyond the TNR unit was established at 2.84 μM dNTPs for the miniforks containing 5′CTG (Figure 6(a), lane 4) and 5′CAG (Figure 6(a), lane 10) repeats. The results showed that the DNA synthesis profiles prior to the TNR unit were very similar for both miniforks (Figure 6(b)) indicating that the sequence downstream of the TNR unit did not influence the activity of the reconstituted replication complex. In contrast, a strong blockage of DNA synthesis was observed as soon as the gp41-gp43-gp59 replication trio hit the 5′CTG sequence (Figure 6(c), left panel). This strong block was specific to the minifork containing the 5′CTG sequence as a smooth profile of DNA synthesis was observed with the minifork containing the 5′CAG sequence (Figure 6(c), right panel). These results showed that at low concentrations of dNTPs, the 5′CTG repeat created a greater impediment to the minimal reconstituted T4 replication complex than the 5′CAG repeat. Consequently, the situation at low concentrations of dNTPs differed from that observed at high concentrations of dNTPs, because at high concentrations of dNTPs, the 5′CAG repeat created a greater impediment to the minimal reconstituted T4 replication complex than the 5′CTG repeat (Figures 4(B)–4(D)). All together, our data suggest that the steps that control the DNA synthesis reaction at low and high concentrations of dNTPs are different.

3.7. Incorporation of dAMP Across TMP in the 5′CTG Sequence Context Is a More Difficult Reaction than the Incorporation of TMP across dAMP in the 5′CAG Sequence Context. At low concentration of dNTPs, the 5′CTG repeat located on the leading strand template created a greater impediment to the reconstituted gp41-gp43-gp59 replication complex than the 5′CAG repeat. To investigate which dNMP incorporation reaction was responsible for the impediment of the DNA polymerase, a coupled leading strand DNA synthesis assay was performed with one of the four dNTPs at a low concentration. The three other dNTPs were kept at a high concentration. We first established the DNA synthesis profiles of the gp41-gp43-gp59 replication trio across a ds random sequence. As shown in Figure 7(a), for a minifork containing a random sequence, the DNA synthesis profile was specific to each reaction condition. For instance, the reaction performed at low concentration of dATP (Figure 7(a), lane 2) gave shorter DNA synthesis products than the reaction performed at low concentration of dGTP (Figure 7(a), lane 5). To quantify this aspect of the reaction, we counted the number of each dNMP to be incorporated up to the TNR insertion site (indicated by a backward arrow in Figure 7(a)) and measured the amount of DNA that was synthesized past the TNR insertion site under each of the four reaction conditions (low [dATP], low [TTP], low [dCTP], or low [dGTP]). To facilitate comparisons of efficiency of DNA synthesis between the miniforks, the amount of DNA synthesized under a given condition was calculated relative to that formed under low concentration of dGTP. There was a clear inverse correlation between the number of dXMP to be incorporated up to the insertion site and the quantity of DNA that was synthesized past the TNR insertion site when this dXTP was present at a low concentration (Figure 7(b)). For instance, 20 dAMP and 9 dGMP had to be incorporated up to the insertion site, and the amount of DNA synthesized beyond the insertion site at low dATP concentration was 16 +/− 5% of that formed at low concentration of dGTP. The same inverse correlation applied to the minimal reconstituted T4 replisome synthesizing through the non structure-forming 5′GTT TNR sequence (Figure 7(c)). In the case of a minifork containing a given TNR sequence, the number of dXMP to be incorporated up to the end of the TNR unit and the amount of DNA synthesized beyond the TNR unit when this dXTP was present at a low concentration were compared. Surprisingly, this inverse correlation no longer held when the gp41-gp43-gp59 replication trio replicated through a structure-forming TNR (5′CAG or 5′CTG, Figures 7(d) and 7(e)). For instance, the incorporation of dGMP opposite dCMP of the 5′CAG or 5′CTG sequence gave the highest yield of DNA synthesis although the number of dGMP to be incorporated up to the end of the TNR unit (32 in case of the 5′CAG containing minifork and 27 in case of the 5′CTG containing minifork) was not the lowest among the four dNMP to be incorporated. In addition, for both miniforks containing a structure-forming TNR, the same

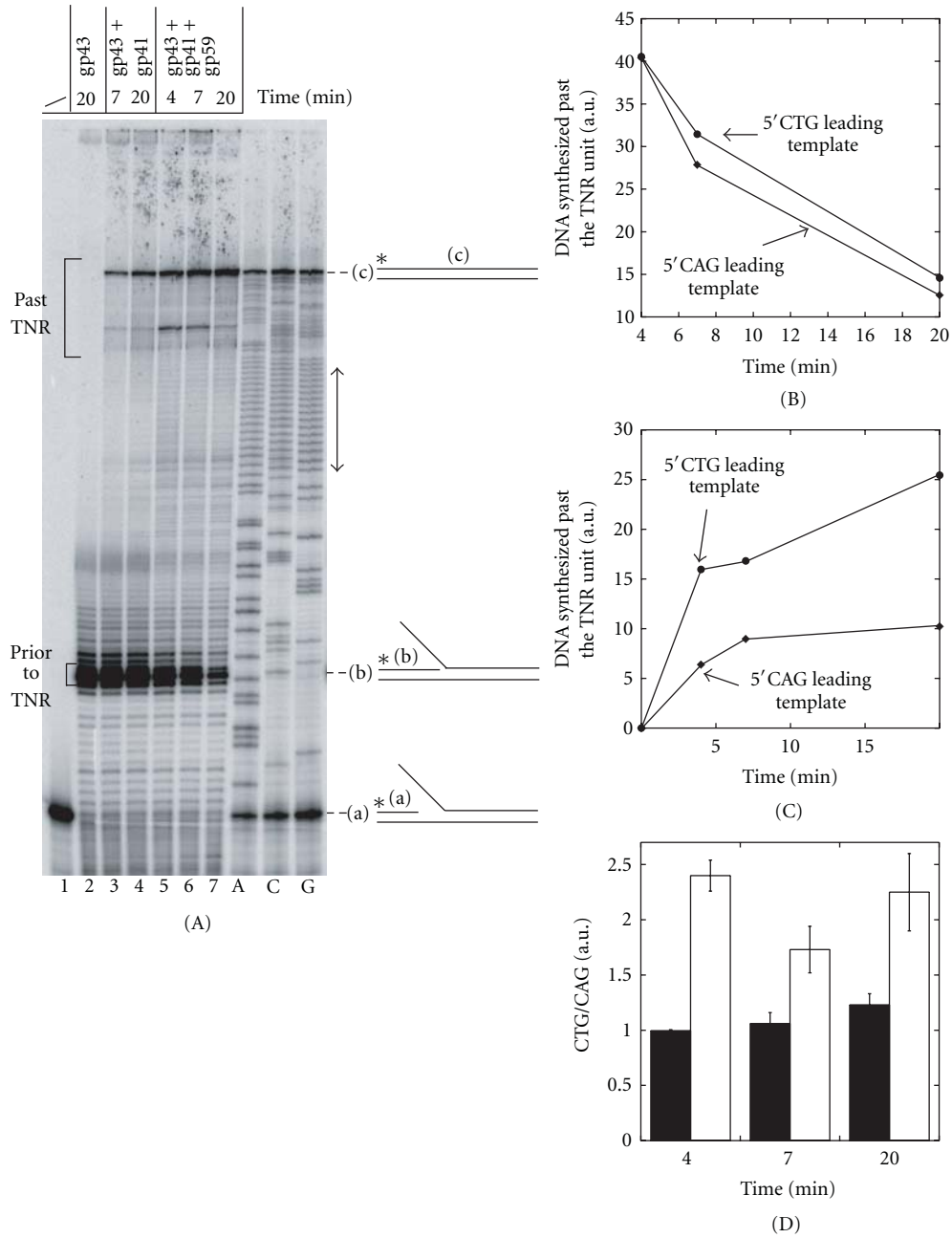


Figure 4: At high concentrations of dNTPs, 5′CAG repeats create a greater impediment than 5′CTG repeats to the gp41-gp43-gp59 replication trio. (A) The leading strand template of the minifork carries 17 repeats of 5′CTG. Three of the four sequencing reactions (A, C, G) of the leading strand template of the minifork are shown on the right side of the figure. The TNR unit is designated by a double-headed arrow. The minifork was incubated with gp43 alone (lane 2), with gp43 and gp41 (lanes 3 and 4), or with gp43, gp41 and gp59 (lanes 5–7). The reaction was quenched at various times and the samples were loaded onto a denaturing sequencing gel. (a), (b), and (c) are as in Figure 3. The primers extended by 14 to 16 nts are shown in the bracket labelled "prior to TNR" and are used to estimate the rate of the gp41-gp43-gp59 replication trio before the TNR unit. The elongation products that contain the full-length TNR are shown in the bracket labelled "past TNR" and are used to estimate the rate of the minimal reconstituted T4 replisome across the TNR unit. (B) The graph shows the decrease of intensity of primers elongated by 14 to 16 nts (indicated by the bracket labelled "prior to TNR" in 4) as a function of time for the miniforks containing 5′CTG or 5′CAG repeats on their leading strand template. (C) The graph shows the accumulation of DNA synthesis products that contain the full-length TNR unit (indicated as a bracket labelled "past TNR" in A) as a function of time for the miniforks containing 5′CTG or 5′CAG repeats on their leading strand template. (D) Black bars: ratio between the amount of primers extended by 14 to 16 nts measured with miniforks containing 5′CTG repeats and the amount of primers extended by 14 to 16 nts measured with the miniforks containing 5′CAG repeats at different times. White bars: ratio between the amount of DNA synthesis products that contain the full-length CAG unit and the amount of DNA synthesis products that contain the full-length CTG unit at different times. The error bars correspond to the standard deviation calculated from at least two independent experiments. B and C were obtained from quantifying the experiment presented in 4(A), and the experiment was performed in parallel with a minifork containing 23 5′CAG repeats. a.u.: arbitrary unit.

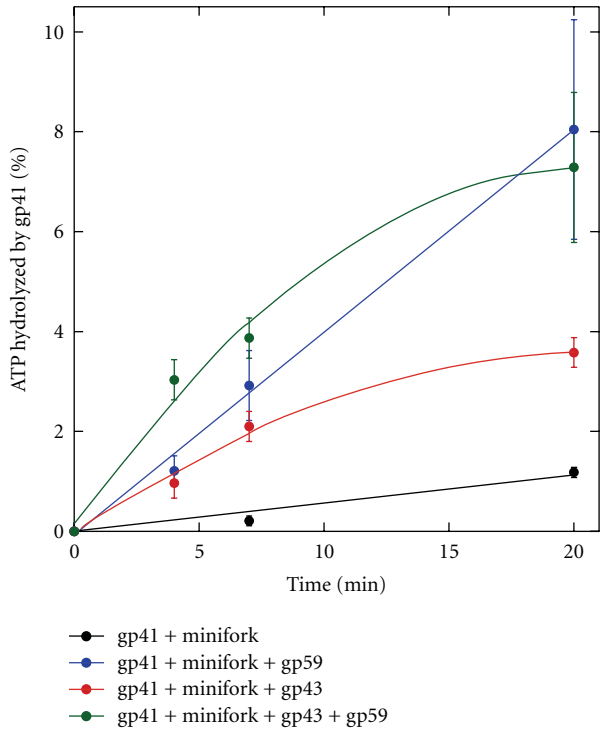


Figure 5: % ATP hydrolyzed by gp41 as a function of time in the presence of the minifork containing 17 repeats of 5′CTG and various proteins. The leading strand template of the minifork carried 17 repeats of 5′CTG and was included in all assays. The gp41 ATPase activity was measured as a function of time under various conditions: black filled circles and black curve (gp41 + minifork); blue filled circles and blue curve (gp41 + minifork + gp59); red filled circles and red curve (gp41 + minifork + gp43); green filled circles and green curve (gp41 + minifork + gp43 + gp59).

amount of dAMP or TMP [37] had to be incorporated up to the end of the 5′CTG or 5′CAG unit, respectively (Figures 7(d) and 7(e)), but the relative amount of DNA synthesized past the 5′CAG unit under low concentration of TTP (40 +/− 13%) was roughly twice that synthesized past the 5′CTG unit under low concentration of dATP (22 +/− 6%) (Figures 7(d) and 7(e)). This result suggested that the incorporation of dAMP across TMP in the 5′CTG sequence context was a more difficult reaction than the incorporation of TMP across dAMP in the 5′CAG sequence context, thus giving an explanation for the greater impediment to the gp41-gp43-gp59 replication trio created by a 5′CTG unit than a 5′CAG unit at low dNTP concentrations.

4. Discussion

Repetitive sequences, including TNRs, are more prone to mutate than random sequences Ellegren [1]. Recently, we proposed a new model (called the template-push model) for the dependence of TNR instability on the orientation of the replication fork and the deletion bias observed *in vivo* for these repetitive sequences [26]. In this model, the TNR sequence that the replisome must replicate creates a greater hindrance for the progression of the leading than the lagging DNA polymerase; as a consequence, the replicative helicase and the leading DNA polymerase transiently uncouple their activities, and a short gap of ss DNA between the two proteins appears. To restore its coupling with the moving helicase and save time for DNA synthesis, the leading DNA polymerase passes over the small track of naked leading strand template without synthesizing DNA. By this mechanism, polymerase-helicase coupling is maintained but at the expense of a hairpin that is formed on the template strand after protein coupling has been re-established. If it is left unrepaired or is repaired in an error-prone manner, the hairpin can induce a deletion of the TNR unit at the next round of replication. In this paper, we describe a method to build replication miniforks suitable for testing the template-push model, and we show how the use of these miniforks brought insights into the mechanism of TNR instability.

Although very simple to prepare, the miniforks assembled from purchased oligonucleotides are limited in size by the length of the oligonucleotides that can be chemically synthesized (around 100 nts). Our method of preparation of replication miniforks overcomes this limitation since the production of ss leading and lagging strand templates relies on PCR followed by the specific degradation of one strand of the PCR product by the T7 exonuclease (Figure 1). As a consequence any sequence carried in a ds DNA can be used, making it possible to assemble a replication minifork of any given sequence. The Herculase-enhanced DNA polymerase and the Phusion high-fidelity DNA polymerase have both been successfully used to amplify sequences containing around 20 TNRs. The leading strand template indeed carries the expected number of repeats (see sequencing lanes in Figures 3, 4, 6, and 7). It is possible that the Herculase-enhanced DNA polymerase becomes more appropriate than the Phusion high-fidelity DNA polymerase when dealing with longer repeats because the former DNA polymerase can faithfully and efficiently cope with long G-C rich targets (Stratagene).

The striking conservation of the DNA replication apparatus in bacteriophage T4 and in human cells [27] make the T4 DNA replication machinery an ideal simple model system to test *in vitro* the miniforks prepared by the method described above and investigate the activity of reconstituted replisomes of increasing complexities. A functional minimal replication complex composed of the T4 DNA polymerase and the T4 helicase can perform strand displacement DNA synthesis across a random sequence, nonstructure-forming and structure-forming TNRs (Figures 3 and 4). The T4 helicase loader stimulates both the DNA synthesis and the ATPase activities of the helicase-DNA polymerase replication couple, but interestingly to different extents. For instance, the twofold stimulation of the ATPase activity of the T4 helicase by the T4 helicase loader (Figure 5) is associated with the fivefold increase of leading strand DNA synthesis (Figure 4). As indicated by its name, the T4 helicase loader stimulates the loading of the T4 helicase around naked or SSB-covered ss DNA. Whether this factor dissociates after loading of the helicase is unclear, and it is possible that the T4 helicase loader remains part of the T4 replisome. If

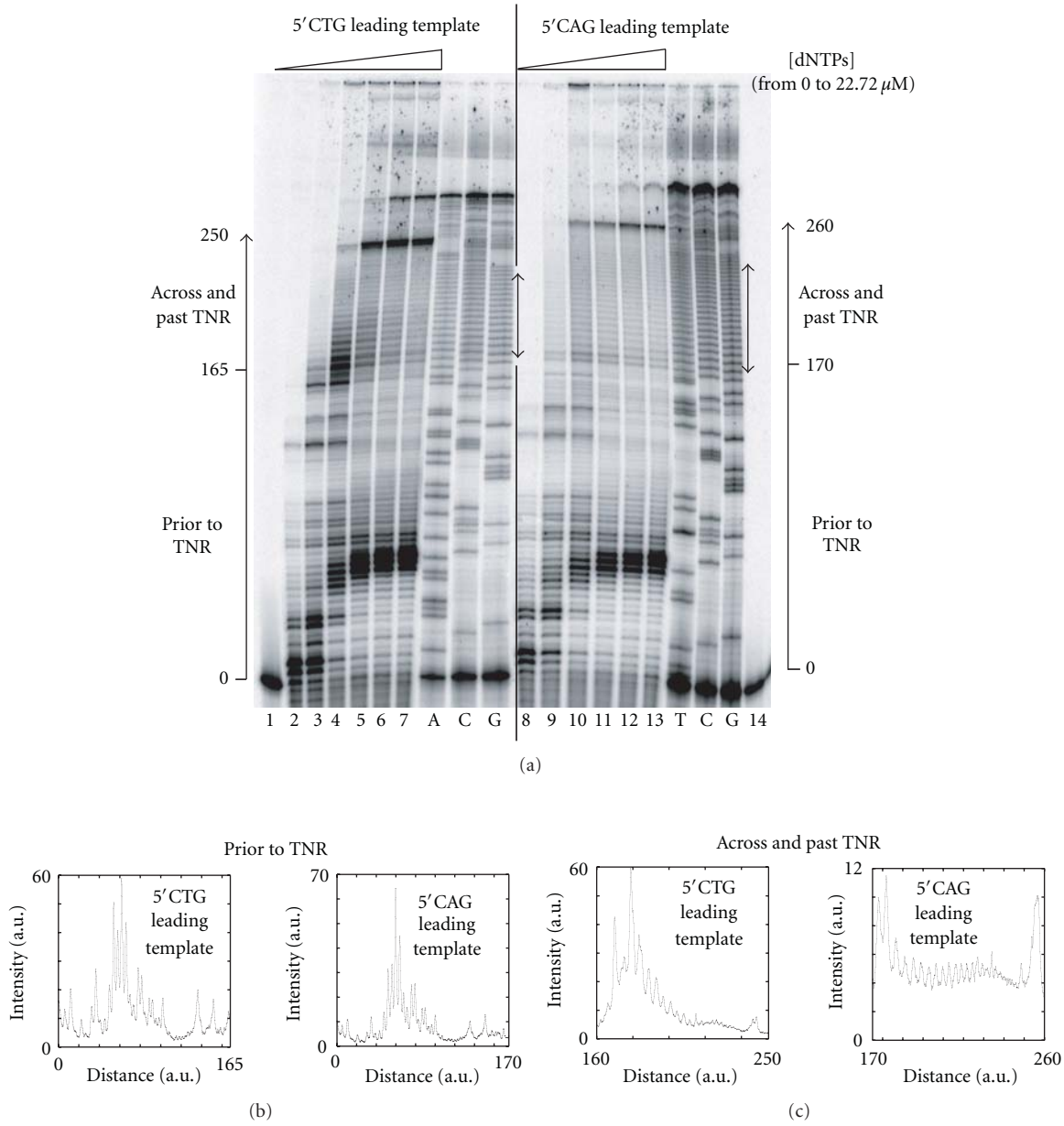


Figure 6: At low concentrations of dNTPs, 5′CTG repeats create a greater impediment than 5′CAG repeats to the gp41-gp43-gp59 replication trio. (a) Miniforks containing 17 5′CTG repeats (left panel) and 23 5′CAG repeats (right panel) on the leading strand template were incubated with gp43, gp41, and gp59 for 10 min at various concentrations of dNTPs, and the reaction products were resolved on a denaturing sequencing gel. Each of the four dNTPs is present at the following concentration (in μM): 0 (lanes 1 and 14); 0.71 (lanes 2 and 8); 1.42 (lanes 3 and 9); 2.84 (lanes 4 and 10); 5.68 (lanes 5 and 11); 11.36 (lanes 6 and 12); 22.72 (lanes 7 and 13). Three of the four sequencing reactions (A, C, G for the minifork containing the 5′CTG repeats, left; T, C, G for the minifork containing the 5′CAG repeats, right) of the leading strand template of the minifork are shown on the right side of the dNTP titrations. The TNR unit is designated by a double-headed arrow. The lines along which DNA synthesis profiles at 2.84 μM dNTPs (lanes 4 and 10) have been established are shown on the side of each panel and are divided into two parts labelled "prior to TNR" (from position 0 to 165 for the miniforks containing the 5′CTG repeats or position 0 to 170 for the miniforks containing the 5′CAG repeats) and "across and past TNR" (from position 160 to 250 for the miniforks containing the 5′CTG repeats or position 170 to 260 for the miniforks containing the 5′CAG repeats). (b) DNA synthesis profile at 2.84 μM dNTPs prior to the TNR unit from position 0 to 165 for the miniforks containing the 5′CTG repeat (left panel) and from position 0 to 170 for the miniforks containing the 5′CAG repeats (right panel). (c) DNA synthesis profile at 2.84 μM dNTPs across and beyond the TNR unit from position 160 to 250 for the miniforks containing the 5′CTG repeat (left panel) and from position 170 to 260 for the miniforks containing the 5′CAG repeats (right panel). a.u.: arbitrary unit.

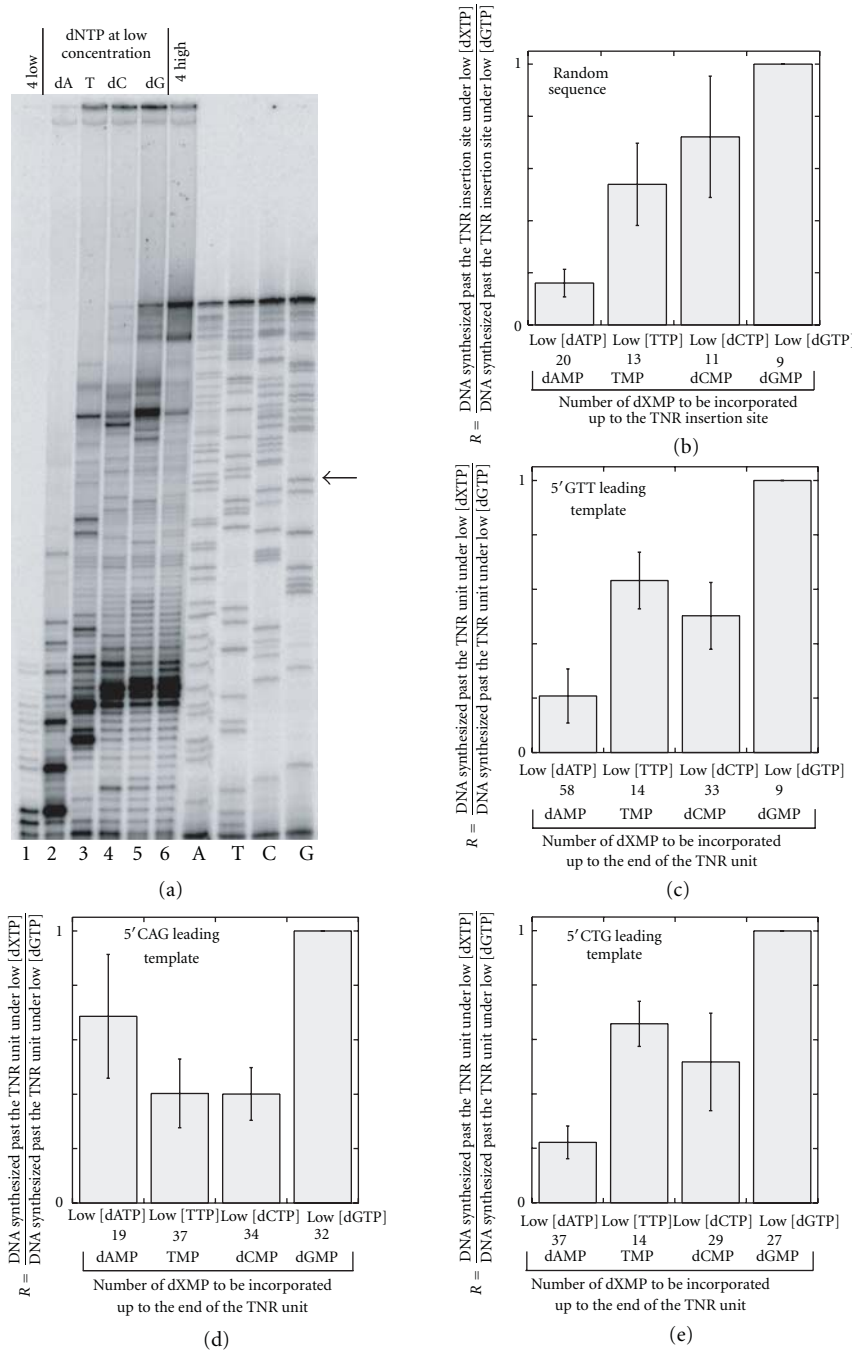


Figure 7: DNA synthesis profiles and efficiencies at various concentrations of dNTPs. (a) The minifork used here carries a random sequence. The four sequencing reactions (A, T, C, and G) of the leading strand template of the minifork are shown on the right side of the gel. The minifork was incubated with gp43, gp41 and gp59 for 10 min with the indicated dNTP mixture and the samples were loaded onto a denaturing sequencing gel. "4 low" (lane 1) means that the 4 dNTPs are at a low concentration (1 μM). "4 high" (lane 6) means that the 4 dNTPs are at a high concentration (80 μM). The single dNTP at low concentration (1 μM) is indicated on the top of the figure (dATP, TTP, dCTP, and dGTP in lanes 2, 3, 4, and 5, resp.). The three other dNTPs are at 80 μM. The backward arrow points to the TNR insertion site. (b) The gel shown in 7(a) has been quantified and the ratio R is shown for each reaction condition (low [dATP], low [TTP], low [dCTP], and low [dGTP]). R is the ratio between the amount of DNA synthesized past the TNR insertion site under a given condition of [dNTPs] (low [dATP] (lane 2 of Figure 7(a)), low [TTP] (lane 3 of Figure 7(a)), low [dCTP] (lane 4 of Figure 7(a)), or low [dGTP] (lane 5 of Figure 7(a)) and the amount of DNA synthesized past the TNR insertion site at low [dGTP] (lane 5 of Figure 7(a)). The number of each dNTP that has to be incorporated up to the insertion site is also indicated underneath the panel of histograms. The same quantification as that described in 7(b) was performed with miniforks carrying a 5′GTT (c), 5′CAG (d) and 5′CTG (e) repeat. For the miniforks containing a TNR unit, R is the ratio between the amount of DNA synthesized past the TNR unit under a given condition of [dNTPs] (low [dATP], low [TTP], low [dCTP], or low [dGTP]) and the amount of DNA synthesized past the TNR unit at low [dGTP]. The number of each dNTP that has to be incorporated up to the end of the TNR unit is indicated underneath the panel of histograms.

the T4 helicase loader travels with the replication complex and if most of the ATP is used to unwind the parental ds DNA (which we believe is true as under our experimental conditions, gp59 stimulates both the intrinsic and the ss-dependent ATPase activity of gp41), our result suggests that the T4 helicase loader prevents the slippage of T4 helicase, by reducing the amount of ATP hydrolyzed that is not associated with forward translocation.

Using the method that quantifies the extent of impediment created by various TNR sequences to the progression of DNA polymerases [26] and by applying it to minimal reconstituted T4 replication complexes, we found that at high concentrations of dNTPs a 5′CAG leading strand template creates a greater impediment to the gp41-gp43-gp59 replication trio than a 5′CTG leading strand template (Figure 4). A similar ranking of these two sequences was reported when naked 5′CAG and 5′CTG sequences were tested in a primer extension assay that did not required the T4 helicase [26]. This result suggests that the ss DNA exposed either by the T4 helicase or by a chemical denaturing treatment is similarly converted into ds DNA by the T4 DNA polymerase. In contrast, when the concentrations of all dNTPs are low, a 5′CTG leading strand sequence more dramatically hinders the progression of the gp41-gp43-gp59 replication trio than a 5′CAG leading strand sequence (Figure 6), suggesting a change in the rate-limiting step when the concentration of dNTPs varies. It is possible that at low dNTP concentrations, the binding of the incoming complementary dNTP by the DNA polymerase becomes the rate-limiting step of the coupled DNA synthesis reaction, whereas at high dNTP concentrations another step, such as the chemical incorporation of dNMP or the translocation of the DNA polymerase after the dNMP incorporation, becomes rate limiting.

By keeping a single (out of the four) dNTPs at a low concentration, it was found that the incorporation of dAMP across TMP of the 5′CTG sequence is a more difficult reaction than the incorporation of TMP across dAMP of the 5′CAG sequence context (Figure 7), thus giving an explanation for the greater hindrance to progression of the minimal reconstituted T4 replisome created by a 5′CTG sequence than a 5′CAG sequence under low concentration of dNTPs (Figure 6). It is well known that dNTPs are not all four present at the same concentration in the cell, TTP being the most abundant dNTP [28–30], and that specific imbalanced pools of dNTPs can be mutagenic [31]. In *E. coli*, the pool of dNTPs drops during cell growth and when the cell transits from exponential growth to stationary phase [29]. Similarly, confluent human cells have a low pool of dNTPs [30]. In addition, in eukaryotic cells, the pool of dNTPs is tightly regulated during the cell cycle and peaks during the S, G2 and M phases [28, 31–33]. During the G1 phase, the amount of dNTPs is low and the use of dNTPs is restricted to mitochondrial DNA synthesis and repair. The fact that the concentration of dNTPs can regulate the progression of the gp41-gp43-gp59 replication trio across a 5′CAG/5′CTG TNR unit in a strand-dependent manner suggests that the repair of a lesion or a gap located in a 5′CAG/5′CTG sequence context, can have different outcomes if it takes place in the G1 or outside the G1 phase of the cell cycle. For instance, during base excision repair, after removal of a 2-hydroxyadenine located in a 5′CAG sequence context, dAMP needs to be incorporated opposite TMP of the CTG repeat. The one-nt gap intermediate formed after the excision of the lesion may have more chance to slip and to give rise to frameshift mutations during the G1 phase than early in the S phase before the passage of the replisome or during the G2 phase. At this stage of the cell cycle, the pool of dNTPs is indeed low and, as pointed out by our experiments, incorporation of dAMP across TMP is difficult. Our results therefore point to a specific role of the phase of the cell cycle (G1 *versus* G2 or S) and the state of the cell (dividing *versus* nondividing) at which DNA lesions are repaired.

Abbreviations

ATP:	Adenosine triphosphate
bp:	Base pair
ds:	Double-stranded
dNTP:	Deoxynucleoside triphosphate
dNMP:	Deoxynucleoside monophosphate
DTT:	Dithiothreitol
EDTA:	Ethylenediaminetetraacetic acid
EtBr:	Ethidium bromide
gp:	gene product
nt:	Nucleotide
PCR:	Polymerase chain reaction
PEI:	Polyethylene imine
PNK:	Polynucleotide kinase
p-t:	Primer-template
1X TBE:	90 mM Tris, 90 mM Borate and 2 mM EDTA
TE:	10 mM Tris-HCl pH 7.8 and 1 mM EDTA
TLC:	Thin layer chromatography
TNR:	Trinucleotide repeat
ss:	Single-stranded
SSB:	Single-stranded binding.

Acknowledgments

This work was supported by the CNRS, Institut Curie, Institut Jacques Monod and Université 7 Paris Diderot. The authors thank Giuseppe Villani and Anne-Lise Haenni for critical reading of the paper.

References

[1] H. Ellegren, "Microsatellite mutations in the germline: implications for evolutionary inference," *Trends in Genetics*, vol. 16, no. 12, pp. 551–558, 2000.

[2] R. Galant and S. B. Carroll, "Evolution of a transcriptional repression domain in an insect Hox protein," *Nature*, vol. 415, no. 6874, pp. 910–913, 2002.

[3] H. P. Gerber, K. Seipel, O. Georgiev et al., "Transcriptional activation modulated by homopolymeric glutamine and proline stretches," *Science*, vol. 263, no. 5148, pp. 808–811, 1994.

[4] M. D. Vinces, M. Legendre, M. Caldara, M. Hagihara, and K. J. Verstrepen, "Unstable tandem repeats in promoters confer transcriptional evolvability," *Science*, vol. 324, no. 5931, pp. 1213–1216, 2009.

[5] J. W. Fondon III and H. R. Garner, "Molecular origins of rapid and continuous morphological evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 52, pp. 18058–18063, 2004.

[6] E. A. D. Hammock and L. J. Young, "Genetics: microsatellite instability generates diversity in brain and sociobehavioral traits," *Science*, vol. 308, no. 5728, pp. 1630–1634, 2005.

[7] L. B. Li and N. M. Bonini, "Roles of trinucleotide-repeat RNA in neurological disease and degeneration," *Trends in Neurosciences*, vol. 33, no. 6, pp. 292–298, 2010.

[8] C. T. McMurray, "Mechanisms of trinucleotide repeat instability during human development," *Nature Reviews Genetics*, vol. 11, no. 11, pp. 786–799, 2010.

[9] E. Roze, C. Bonnet, S. Betuing, and J. Caboche, "Huntington's disease," *Advances in Experimental Medicine and Biology*, vol. 685, pp. 45–63, 2010.

[10] R. Santos, S. Lefevre, D. Sliwa, A. Seguin, J. M. Camadro, and E. Lesuisse, "Friedreich ataxia: molecular mechanisms, redox considerations, and therapeutic opportunities," *Antioxidants and Redox Signaling*, vol. 13, no. 5, pp. 651–690, 2010.

[11] P. K. Todd and H. L. Paulson, "RNA-mediated neurodegeneration in repeat expansion disorders," *Annals of Neurology*, vol. 67, no. 3, pp. 291–300, 2010.

[12] A. Duval and R. Hamelin, "Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability," *Cancer Research*, vol. 62, no. 9, pp. 2447–2454, 2002.

[13] A. Duval and R. Hamelin, "Réparation des erreurs de réplication, microsatellites et cancer," *Medecine/Sciences*, vol. 19, no. 1, pp. 55–62, 2003.

[14] G. Liu, X. Chen, J. J. Bissler, R. R. Sinden, and M. Leffak, "Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells," *Nature Chemical Biology*, vol. 6, no. 9, pp. 652–659, 2010.

[15] M. C. Young, M. K. Reddy, and P. H. von Hippel, "Structure and function of the bacteriophage T4 DNA polymerase holoenzyme," *Biochemistry*, vol. 31, no. 37, pp. 8675–8690, 1992.

[16] N. G. Nossal, "Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model," *Journal of the Federation of American Societies for Experimental Biology*, vol. 6, no. 3, pp. 871–878, 1992.

[17] S. J. Benkovic, A. M. Valentine, and F. Salinas, "Replisome-mediated DNA replication," *Annual Review of Biochemistry*, vol. 70, pp. 181–208, 2001.

[18] K. N. Kreuzer and J. R. Brister, "Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the virology journal series on bacteriophage T4 and its relatives," *Virology Journal*, p. 358, 2010.

[19] T. C. Mueser, J. M. Hinerman, J. M. Devos, R. A. Boyer, and K. J. Williams, "Structural analysis of bacteriophage T4 DNA replication: a review in the virology journal series on bacteriophage T4 and its relatives," *Virology Journal*, vol. 7, p. 359, 2010.

[20] S. K. Perumal, H. Yue, Z. Hu, M. M. Spiering, and S. J. Benkovic, "Single-molecule studies of DNA replisome function," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1094–1112, 2010.

[21] C. F. Morris, H. Hama-Inaba, D. Mace, N. K. Sinha, and B. Alberts, "Purification of the gene 43, 44, 45, and 62 proteins of the bacteriophage T4 DNA replication apparatus," *Journal of Biological Chemistry*, vol. 254, no. 14, pp. 6787–6796, 1979.

[22] C. F. Morris, L. A. Moran, and B. M. Alberts, "Purification of gene 41 protein of bacteriophage T4," *Journal of Biological Chemistry*, vol. 254, no. 14, pp. 6797–6802, 1979.

[23] R. W. Richardson and N. G. Nossal, "Characterization of the bacteriophage T4 gene 41 DNA helicase," *Journal of Biological Chemistry*, vol. 264, no. 8, pp. 4725–4731, 1989.

[24] S. W. Morrical, K. Hempstead, and M. D. Morrical, "The gene 59 protein of bacteriophage T4 modulates the intrinsic and single-stranded DNA-stimulated ATPpase activities of gene 41 protein, the T4 replicative DNA helicase," *Journal of Biological Chemistry*, vol. 269, no. 52, pp. 33069–33081, 1994.

[25] N. G. Nossal, D. M. Hinton, L. J. Hobbs, and P. Spacciapoli, "Purification of bacteriophage T4 DNA replication proteins," *Methods in Enzymology*, vol. 262, pp. 560–584, 1995.

[26] E. Delagoutte, G. M. Goellner, J. Guo, G. Baldacci, and C. T. McMurray, "Single-stranded DNA-binding protein in vitro eliminates the orientation-dependent impediment to polymerase passage on CAG/CTG repeats," *Journal of Biological Chemistry*, vol. 283, no. 19, pp. 13341–13356, 2008.

[27] H. Bernstein and C. Bernstein, "Bacteriophage T4 genetic homologies with bacteria and eucaryotes," *Journal of Bacteriology*, vol. 171, no. 5, pp. 2265–2270, 1989.

[28] G. Bray and T. P. Brent, "Deoxyribonucleoside 5'-triphosphate pool fluctuations during the mammalian cell cycle," *Biochimica et Biophysica Acta*, vol. 269, pp. 184–191, 1972.

[29] M. H. Buckstein, J. He, and H. Rubin, "Characterization of nucleotide pools as a function of physiological state in Escherichia coli," *Journal of Bacteriology*, vol. 190, no. 2, pp. 718–726, 2008.

[30] P. Ferraro, E. Franzolin, G. Pontarin, P. Reichard, and V. Bianchi, "Quantitation of cellular deoxynucleoside triphosphates," *Nucleic Acids Research*, vol. 38, no. 6, p. e85, 2010.

[31] D. Kumar, J. Viberg, A. K. Nilsson, and A. Chabes, "Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint," *Nucleic Acids Research*, vol. 38, no. 12, pp. 3975–3983, 2010.

[32] A. Chabes, B. Georgieva, V. Domkin, X. Zhao, R. Rothstein, and L. Thelander, "Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase," *Cell*, vol. 112, no. 3, pp. 391–401, 2003.

[33] A. Chabes and B. Stillman, "Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast Saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 4, pp. 1183–1188, 2007.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 718974, 11 pages
doi:10.4061/2011/718974

Review Article

Arsenic Biotransformation as a Cancer Promoting Factor by Inducing DNA Damage and Disruption of Repair Mechanisms

Victor D. Martinez,[1,2] Emily A. Vucic,[1] Marta Adonis,[2] Lionel Gil,[2] and Wan L. Lam[1]

[1] *Department of Integrative Oncology, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3*
[2] *Biomedical Sciences Institute, Faculty of Medicine, University of Chile, Independencia 1027, 8380453 Santiago, Chile*

Correspondence should be addressed to Victor D. Martinez, vmartinez@bccrc.ca

Received 16 March 2011; Accepted 6 June 2011

Academic Editor: Frédéric Coin

Chronic exposure to arsenic in drinking water poses a major global health concern. Populations exposed to high concentrations of arsenic-contaminated drinking water suffer serious health consequences, including alarming cancer incidence and death rates. Arsenic is biotransformed through sequential addition of methyl groups, acquired from s-adenosylmethionine (SAM). Metabolism of arsenic generates a variety of genotoxic and cytotoxic species, damaging DNA directly and indirectly, through the generation of reactive oxidative species and induction of DNA adducts, strand breaks and cross links, and inhibition of the DNA repair process itself. Since SAM is the methyl group donor used by DNA methyltransferases to maintain normal epigenetic patterns in all human cells, arsenic is also postulated to affect maintenance of normal DNA methylation patterns, chromatin structure, and genomic stability. The biological processes underlying the cancer promoting factors of arsenic metabolism, related to DNA damage and repair, will be discussed here.

1. Introduction

Arsenic is one of the most abundant elements in the Earth's crust [1]. Chemically, it is classified as metalloid, exhibiting organic (when linked with carbon and hydrogen) and inorganic (combined with oxygen, chlorine, and sulfur, among other elements) forms [2]. Inorganic arsenic (iAs) can be present naturally in soil, especially in rocks containing copper or lead, and in the atmosphere as airborne dust. Additionally, anthropogenic activities, such smelter operations, can cause water contamination [3, 4]. In the environment, iAs can be found in several oxidation states, more frequently as trivalent (iAs[III], also known as arsenite) and pentavalent (iAs[V] or arsenate) species [5]. These forms are differently metabolized by mammals (see below) and exhibit distinct grades of toxicity.

Several health effects have been documented as a consequence of iAs exposition, with the majority of harmful exposure coming from ingestion through drinking water. iAs-associated malignancies include skin lesions, hypertension, ischemia, some endemic peripheral vascular disorders (e.g., "black foot disease"), severe arteriosclerosis, neuropathies, noticeably, many types of cancer [6–9]. A number of studies have established significant associations and/or dose response trends between iAs in drinking water and occurrence of tumors of the skin, bladder, kidney, liver, prostate, and lungs [10–15].

The evidence of a relationship between iAs in drinking water and cancer is extensive and sufficient, leading to the International Agency of Research on Cancer (IARC) to consider this metalloid as a Group 1 human carcinogen. The estimated cancer-death risk associated with daily consumption of 1.6 liters of water with iAs concentrations of 50 μg/L is 21/1,000 [16]. For these reasons, the World Health Organization and the U.S. Environmental Protection Agency have recommended a threshold of 10 μg/L for iAs concentration in drinking water [17, 18].

Despite efforts to reduce high-scale exposure, many nations throughout the world have iAs concentrations in water that are above the recommended level [19–21]. Approximately 40 million people worldwide are thought to be exposed to iAs levels that can be considered dangerous

[19]. Among them, 21 million people in Bangladesh and India (West Bengal) are exposed to drinking water with i*As* concentrations >50 μg/L [22], and shockingly, iAs concentration in water wells in these areas has been documented as high as 1000 μg/L [23]. In China, it has been estimated that more than 3 million people are exposed to iAs from groundwater [24], while in southwestern Taiwan, some residents have used well water contaminated with iAs for more than 50 years (some ingesting as much as 1000 μg iAs/day) [25–27]. In Northern Chile, the population was exposed to levels of iAs in drinking water around 900 μg/L between 1958 and 1970, with nearby towns registering exposures of 600 μg/mL as late as 1994 [11].

2. Arsenic Biotransformation

About 80–90% of ingested As[III] or As[V] is absorbed from the gastrointestinal tract [28–30]. Data derived from autopsies has determined that muscles, bones, kidneys, and lungs have the highest absolute accumulated amounts of iAs, while skin and excretory/storage organs, such as nails and hair, are the most concentrated [31]. iAs[III] exhibits a significantly higher biological activity than As[V]; however, effects observed in mammals could be similar, since absorbed As[V] is mostly reduced to As[III] on the initial steps of arsenic metabolism in mammals [32, 33]. Interestingly, there is evidence for interindividual differences in iAs metabolism/excretion in humans and other species [34, 35].

The biotransformation process of iAs occurs via methylation through alternating reduction of As[V] to As[III], and subsequent addition of methyl groups [36]. This methylation process uses S-adenosylmethionine (SAM) as a methyl group donor, through a SAM-dependant As[III]-methyltransferase, initially isolated from rat liver and a human homologue of cytochrome19 [37]. This enzyme catalyzes the transfer of a methyl group from SAM to As[III] producing methylated and dimethylated arsenic compounds. Trivalent methylated species, such monomethylarsonous acid (MMA[III]) and dimethylarsinic acid (DMA[III]), have been detected in the urine of patients chronically exposed to iAs in drinking water [38, 39]. Methylated pentavalent arsenicals such as monomethylarsonic acid (MMA[V]) and dimethylarsinic acid (DMA[V]) are major metabolites of iAs in human urine, with DMA[V] being the final metabolite in humans [39–41]. Derivate methylated species from iAs metabolisms are considered relevant agents during arsenic carcinogenicity, specially through induction of oxidative stress and impairing DNA repair processes. These aspects will be discussed in the following sections.

Despite evidence of biotransformation role in arsenic carcinogenicity, it has been demonstrated that arsenic can induce malignant transformation in cell lines with deficient arsenic-methylation capacity. Arsenic methylation-deficient RWPE-1 human prostate cells undergo malignant transformation when exposed to 5.0 μM of iAs[III] during 30 weeks [42]. Alternative mechanism of arsenic-induced malignant transformation might be associated with mitochondrial dysfunction (see below), specifically through transcription and replication of the mitochondrial genome, in which the mitochondrial transcription factor A (mtTFA) and its regulators, such the nuclear respiratory factor-1 (NRF-1), play key roles [43, 44]. In this context, it has been demonstrated that mtTFA and NRF-1 expressions levels are increased in cells exposed to iAs[III] in a concentration-dependent manner, suggesting that arsenic regulates mitochondrial activity through an NRF-1-dependent pathway [45].

3. Arsenic Carcinogenicity: Role of Oxidative Stress

Despite the strong relationship between iAs exposure and cancer, the exact mechanism is still unknown. There is evidence supporting low level mutagenic activity of iAs; however, it has also been shown that iAs can induce transformation in several cell types [46, 47]. Moreover, iAs can interfere with a number of biological processes, including DNA methylation, since the arsenic biotransformation pathway uses SAM as a methyl group donor. Therefore, epigenetic mechanisms have also been proposed to participate in iAs-induced carcinogenesis [48].

Biotransformation of iAs has been proposed to generate final and intermediate metabolites exhibiting higher toxicity and reactivity compared to originally ingested iAs [5, 49, 50]. Methylated species, especially DMA[V], have been demonstrated to be genotoxic and cytotoxic [46, 49, 51–53]. Several studies have shown that DMA[V] can exhibit carcinogenic potential in mammals, mainly in lungs, skin, liver, kidney, thyroid, and urinary bladder [39, 54–58]. It has been proposed that DMA[V] can participate in promoting tumorigenesis of lungs and skin in mouse via the production of dimethylated arsenic peroxide [$(CH_3)_2AsOO\cdot$], a type of reactive oxygen species (ROS) generated during iAs metabolism [53, 54].

In the light of these facts, oxidative stress has been proposed as a plausible general mode of action for iAs carcinogenesis [59–63]. Oxidative stress is characterized by generation of several ROS, such as superoxide anion (O_2^-), hydroxyl radical ($\cdot OH$), hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$), and peroxyl radical (LOO), among others [64]. One of the primary species formed in iAs-induced oxidative stress is O_2^-, followed by a cascade of secondary ROS such as H_2O_2 and $\cdot OH$ [61].

iAs exposure results in the generation of ROS in various cellular systems, and its production has been proposed as one of the early biological events on iAs-related carcinogenic process [65]. In addition, cultured vascular endothelial cells exposed to iAs increase oxygen cell consumption contributing to increased ROS production, stimulating cell signaling and activating transcription factors [66]. Conversely, ROS scavengers can suppress arsenic-induced oxidative stress and its cytotoxic effects in cells [67, 68]. It has also been described that iAs exposure can affect expression of genes associated with stress-related components, DNA damage and repair-responsive genes, activation of transcription factors such as the AP-1 complex, and increases in proinflammatory

cytokines, which could influence response to acute arsenic toxicity [69]. Alternatively, ROS generation by iAs can involve hepatic and renal heme oxygenase isoform 1, generating among others species, free iron which subsequently participates in reactions where H_2O_2 is reduced to OH^- and ·OH [69]. Additionally, the oxidation of iAs[III] to As[V] during formation of intermediary arsine species can also generate H_2O_2 [70].

Mechanisms of iAs carcinogenicity could vary between different tissues, due to different oxygen concentrations, and accumulation of iAs species, endogenous reducing agents, and ferritin, among others factors [71, 72]. For example, lungs are exposed to the highest oxygen tensions in the body, and DMA[III], and its derivates (including ROS) are excreted through the lung, which could explain why this organ is frequently affected by iAs-induced carcinogens [60].

It has been suggested that arsenic-associated mitochondrial dysfunction, mitochondrial DNA (mtDNA) depletion, and induction of mtDNA deletions may contribute to the carcinogenicity in humans [73]. Also, mitochondria might be an important target of arsenic-induced genotoxicity [74]. On the other hand, since mitochondria is a major source of intracellular ROS, arsenic-mediated disruption of its function can lead to an increase in intracellular ROS levels and subsequently, to an increased mutagenic potential, either directly or by decreasing DNA repair capacity [73]. Relationships between mitochondria and arsenic-mediated effects are supported by observations such as suppression of arsenic-induced apoptosis in HeLa cells by the antioxidant action of N-acetyl-cysteine, which prevents mitochondrial membrane depolarization [75]. Alternatively, arsenic can act directly through condensing mitochondrial matrix and opening of permeability transition pores [76].

4. DNA and Chromosomal Damage by iAS-Induced Oxidative Stress

Genotoxic mechanisms associated with arsenic carcinogenicity remain controversial. While some groups argue against this type of interaction, others have postulated this can be a significant mode of action. Rossman [46] has proposed that arsenite does not react directly with DNA. In the same way, toxic doses (10–15 μM) of iAs[III] act as a poor mutagen at the gpt mutagenic target in transgenic Chinese hamster G12 cells [77]. On the other hand, it has been proposed that iAs[III] is a significant mutagen that induces mainly large chromosomal mutations [78]. Alternatively, arsenic has been shown to be mutagenic to mitochondrial DNA and can potentially induce nuclear DNA damage by activating mitochondrial ROS through increased expression of mtTFA [45]. Also, arsenic can induce mutations as well as methylation changes in the mouse testicular Leydig cell genome [79]. Similarly, comet assay performed on human prostate epithelial cells exposed to 100 pg/mL of arsenic exhibited tail-like structures, suggesting induction of nuclear DNA damage [45].

iAs is known to damage chromosomes [80]. Due to little evidence of covalent binding between iAs and DNA structures, it has been proposed that much of the DNA damage observed during iAs exposure is indirect, occurring mainly as a result of ROS induction which generates DNA adducts, DNA strand breaks, cross links, and chromosomal aberrations [81, 82]. Figure 2 indicates the sequence of events related to ROS induced DNA damage after iAs exposure. Depending on which cell cycle phase exposure occurs, as a consequence DNA oxidation, iAs can result in gross chromosomal aberrations including DNA strand breaks [61, 69].

5. DNA Strand Breaks

iAs can induce DNA strand breaks even at low concentrations. Main related-events are summarized in Figure 1. Single-strand DNA (ssDNA) breaks are the most common lesions induced by exogenous genotoxins [83]. Arsenic-induced ssDNA breaks are likely caused through ROS, either directly by free-radical attack on the DNA bases or indirectly during the course of base excision repair (BER) mechanisms [84]. Arsenic-induced ROS has been shown to promote ssDNA breaks in mice lungs [70]. Furthermore, human fibroblast cell lines exposed to iAs exhibit ssDNA breaks and DNA-protein adducts, as well as sister chromatid exchanges [85].

iAs is also capable of producing double-strand DNA (dsDNA) breaks at concentrations of 5 μM in mammalian cells [86]. These are one of the most deleterious and mutagenic DNA lesions experienced in human cells, leading to gross losses of genetic material [87]. Therefore, iAs is also proposed to act as a cocarcinogen, exacerbating damage induced by other agents. In this context, 1 μM of iAs increases UVR-mediated DNA strand breaks by interfering with Poly-adenosine diphosphate-ribose polymerase 1 (PAPR-1) activity, which plays an important role in the ssDNA or dsDNA breaks repair process [88].

MMA[III] was found to be a potent clastogen in late G1- or S-phase-treated cells; however, lesions induced by MMA[III] are quickly repaired through BER mechanisms when they are induced in G0- or G1-phase of the cell cycle [84]. Trivalent arsenicals might induce either chromatid- or chromosome-type aberrations during treatment in G0 or G1. If ssDNA or dsDNA breaks produced by iAs-induced ROS pass the S-Phase (DNA synthesis), replication occurs and chromatid- and chromosome-type aberrations can be produced [89]. Evidence pertaining to these type of aberrations is discussed below.

6. Arsenic-Induced Chromatid and Chromosomal Aberrations

Arsenic is a known inducer of chromosomal and chromatid aberrations. Lee et al. [90] demonstrated that iAs can effectively induce methotrexate-resistance in mouse 3T6 cells, resulting in selection of cells with amplification of the dihydrofolate reductase gene [91]. Genetic changes were observed in bladder tumor (transitional cell carcinoma, TCC) from 123 patients in Argentina and Chile, exposed to

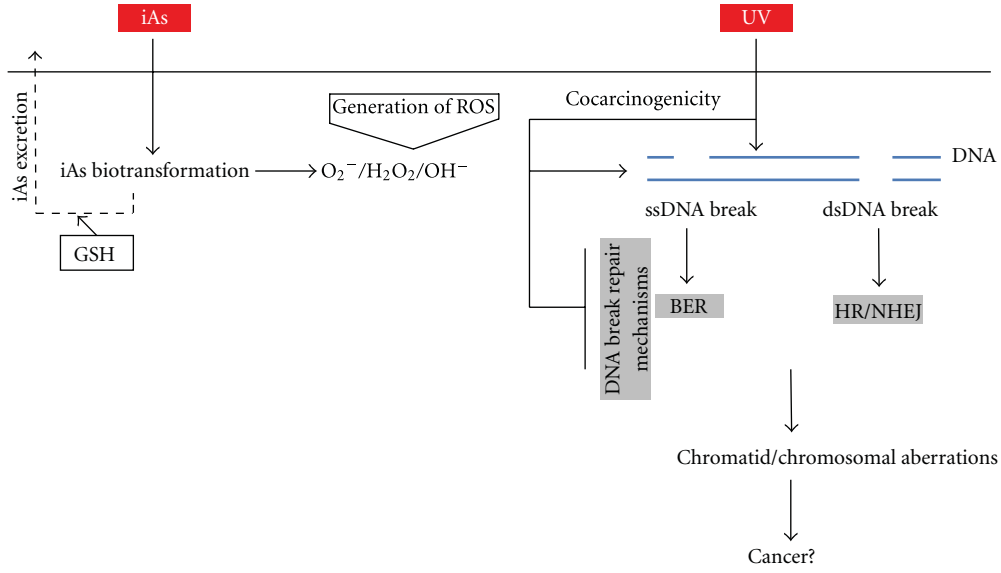


Figure 1: Arsenic-induced DNA strand breaks. After ingestion, iAs biotransformation process could lead to iAs excretion, mainly conjugated with Glutathione (GSH). On the other hand, biotransformation process may generate reactive oxygen species (ROS), probably in a specific sequence: superoxide anions (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($\cdot OH$). These species can induce both single-strand (ssDNA) and double-strand (dsDNA) breaks by inducing oxidative damage. In parallel, they can inhibit DNA break repair mechanisms both for ssDNA breaks (mainly base excision repair [BER]) and for dsDNA breaks (homologous recombination [HR] and/or nonhomologous end joining [NHEJ]). Additionally, ROS derived from iAs biotransformation can act as cocarcinogens, for example, increasing damage potential of ultraviolet (UV) light. All these events could be associated, in part, to iAs-related carcinogenic mechanism.

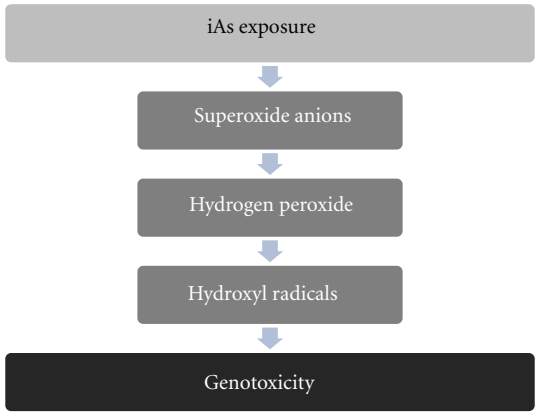


Figure 2: Events related with ROS-induced DNA damage after iAs exposure. Specific sequence of reactive oxygen species generation as a consequence of iAs biotransformation in mammals.

iAs concentrations exceeding 500 μg/L. Individuals exposed to high As concentrations (>300–600 μg/L) exhibited a higher total number of chromosomal aberrations, supporting the hypothesis that exposure to iAs increases genomic instability. Furthermore, chromosomal aberrations (specifically DNA copy-number alterations) were more abundant among iAs-exposed bladder TCC tumors from southwest Taiwan compared with nonexposed tumors from the same area [92]. Some alterations were common to those found in other studies, suggesting that nonrandom events are associated with As-induced urinary TCC formation and progression [93].

Other large-scale cytogenetic aberrations have been observed in iAs-exposed populations. Gonsebatt et al. [94] analyzed cytogenetic effects in individuals exposed to different levels of As in drinking water. People exposed to iAs at an average of 400 μg/L showed a significant increase in frequency of chromatid and isochromatid deletion in first-metaphase lymphocytes and micronuclei in oral and epithelial exfoliated cells compared to individuals with lower exposures. Women and children from the northeast Andean Region of Argentina exposed to 200 μg/L of iAs in drinking water displayed higher micronuclei frequency compared to people exposed to very low iAs concentrations, but did not have altered frequency of other aberrations, such as sister chromatid exchange, specific translocations, or cell-cycle progression [95].

7. Oxidative Damage

DNA modifications due to iAs-induced ROS can produce oxidative damage, which can be measured through the presence in urine of products of guanine oxidation in position 8 (8-oxo-2′-deoxyguanosine (8-oxodG), 8-hydroxyguanine [8-oxo-G], 8-hydroxyguanosine [8-oxy-Guo] and 8-hydroxy-2′-deoxyguanosine [8-OHdG] [64]. After DMA[V] administration in terminal bronchiolar Clara cells from mice, markers for oxidative stress were detected, including 8-oxodG [96]. Additionally, it has been demonstrated that the presence of 8-OHdG was associated with administration

of DMA[V] in iAs-related human keratoses, squamous cell carcinoma, basal cell epithelioma, and normal skin from iAs-intoxicated patients [97–99]. Also, iAs[III] can induce 8-OHdG and promote genomic instability by damaging DNA and inducing oncogene expression (including several factors regulating cell cycle progression) human breast cancer MCF-7 adenocarcinoma epithelial cells exposed to iAs[III] [100]. Oral administration of DMA[V] increases 8-oxo-G levels through $(CH3)_2AsOO\cdot$ [54, 55].

8. Inhibition of DNA Repair Mechanisms Associated with Arsenic Exposure

iAs can also induce DNA damage by interfering with the DNA repair processes. Inhibition or impairment of the DNA repair processes, principally the repair of DNA strand breaks, is considered one of the main mechanisms of iAs carcinogenesis [88, 101, 102]. For example, DMA[V] affects DNA repair and replication mechanisms in human alveolar cells, leading to persistence DNA damage (mainly apurinic/apyrimidinic sites) and generating ssDNA breaks as a consequence [103, 104].

DNA base damage (induced by oxidative stress) can be repaired through excision repair mechanisms, which are subdivided into BER and nucleotide excision repair (NER) pathways [105]. BER is the predominant repair pathway for DNA lesions caused by ROS, and the first candidate in iAs-related DNA repair [69, 106]. Transcription levels of genes related to BER mechanisms are altered in a gene, age-, dose-, and duration-dependent manner in lung tissue of mice exposed to iAs [107]. On the other hand, iAs was also shown to alter BER mechanisms in GM847 lung fibroblasts and HaCaT keratinocytes, increasing levels of BER-related enzymes and repair capacity [108].

Several enzymes participate in the BER mechanism, some of which are known to be modulated by iAs. Among them, DNA polymerase β (Polβ) and DNA ligase I (LIG1) have been described as affected by As[III] [109, 110]. Normally, after generation of 5′ incision on an abasic site leaving a 3′-hydroxyl and a 5′-deoxyribose 5-phosphate, Polβ hydrolyses the 5′-sugar phosphate and adds at least one nucleotide to the 3′-hydroxyl end. The remaining strand is nick sealed by LIG1, and PARP-1 may recruit the required proteins [108]. However, in lung fibroblasts and HaCaT keratinocytes exposed to As[III], Pol β mRNA levels are downregulated in a dose-dependent manner (doses $>1\,\mu$M), and at doses lower than $1\,\mu$M both Pol β mRNA and protein levels, and consequently, BER activity, were significantly increased [108]. Additionally, this enzyme is stimulated in response to low doses iAs and modulated by other sources of oxidative stress [111–114]. Interestingly DNA copy-number alterations (CNAs) in lung squamous cell carcinoma (SqCC) from iAs-exposed patients from northern Chile contain the Polδ 1 (DNA polymerase δ 1, catalytic subunit), which codes for the proofreading domain of the DNA polymerase δ complex and also participates in ssDNA breaks repair process [115–119].

It has been proposed that iAs[III] works at transcriptional level to repress a group of genes encoding for DNA repair enzymes participating in BER and NER mechanisms, mainly through its downregulation. This, in combination with other events, contributes to toxicity or cancer [120]. In parallel, changes in expression levels have been also corroborated in human exposed populations. Exposure to arsenic in drinking water was correlated to decreased expression of ERCC1, XPB, and XPF in lymphocytes from exposed individuals [121]. Decreased ERCC1 gene expression was confirmed in lymphocytes treated with $> 1\,\mu$M of iAs[III], and a significant reduction of ERCC1 protein levels was observed among individuals exposed to drinking water with low levels of arsenic [122]. Similarly, mRNA levels of ERCC1 expression were significantly associated with arsenic concentrations in drinking water, implicating the DNA repair response was induced by arsenic exposure [123]. On the other hand, OGG1 expression (which encodes for 8-oxoguanine DNA glycosylase, involved in base excision repair of 8-oxoguanine [124]) was strongly associated with arsenic concentrations [125], revealing involvement of mechanisms related the effects of arsenic-mediated ROS on DNA.

DNA ligation is a key step in DNA repair pathways [126]. Interestingly, it has been shown that iAs can specifically

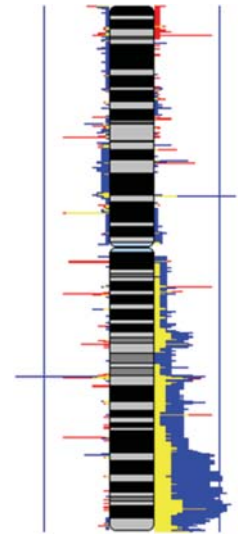


Figure 3: Comparison of CNA frequency at chromosome 3 between lung SqCC exposed and nonexposed to iAs. The figure represents a comparison of CNA frequency at chromosome 3 generated from 52 lung SqCC biopsies by using a submegabase resolution tiling-set rearray (SMRTr) platform. Of those, 22 derived from arsenic-exposed smokers and never smokers patients from Northern Chile (red) and 30 were current and ex-smokers North American patients without known arsenic exposure nonexposed (blue). Frequency of alteration results for exposed and nonexposed SqCCs cases has been overlaid in this figure, with regions in yellow, denoting a sector of overlapping alteration status in both groups. The magnitude of red, yellow and blue bars represents percentage of samples exhibiting corresponding alteration (0–100%, with blue vertical lines representing 50% frequency). DNA gains and losses are represented to the right and left of chromosome, respectively. Adapted from Martinez et al. [115].

inhibit this process as well. More recently we have found that the mRNA, protein and activity levels of both DNA ligase I and ligase III are significantly reduced in mammalian cells in response to As[III] [109]. Additionally, As[III] retards DNA break rejoining by interacting with the vicinal dithiols and thus inhibiting DNA ligation [127]. Mammalian cells have been shown to exhibit a dose-dependent decrease in ligase activity with exposure to As[III], corresponding to a decrease in mRNA levels of this enzyme [108, 128]. On the other hand, it has been also shown that LIG1 and other DNA damage/repair genes were increased by As[III] and As[V] treatment, suggesting a cellular response to iAs-induced DNA damage [129].

Members of the poly (ADP-ribose) polymerase (PARP) family also play an important role in the regulation of DNA damage repair. PARP-1 (accounting for about 90% of the total cellular poly ADP-ribose formation) acts as a "DNA damage sensor", exhibiting high affinity to bind both ssDNA and dsDNA breaks [130–132]. It has been proposed that lack of PARP-1 enhances cellular sensitivity to As[III] [133]. Cells deficient in this gene product display greater telomere attrition. This process can be attributable to susceptibility of the triple-G-containing structures of telomeric DNA to oxidative damage [134, 135]. In parallel, cells deficient in PARP-1 exhibit reduced repair of 8-oxoguanine, another marker for oxidative damage that can potentially be induced by iAs [136]. Finally, specific CNAs located at located at 10q11.23 in lung SqCC from iAs-exposed patients from northern Chile contain the PARG (polyADP-ribose glycohydrolase) gene, which also participates in ssDNA breaks repair process [102, 115, 137].

9. Genomic Landscape of Arsenic-Related Lung Cancer

Lungs are the most frequently affected organ by iAs, and lung cancer remains the main cause of iAs-related death [24]. Tobacco exposure is the main aetiological factor in lung cancer; however, iAs ingestion through drinking water also represents a risk factor, particularly for lung squamous cell carcinomas (SqCCs). Interestingly, the incidence of SqCC is decreasing worldwide and is usually associated with cigarette smoking, but in Northern Chilean populations exposed to arsenic contaminated drinking water, SqCC frequently occurs in never smokers, [14, 138] suggesting distinct molecular tumorigenic pathways may underlie arsenic-related cancers.

To this effect, it was determined if globally, there existed CNAs specific to lung SqCC cases from a Northern Chilean population chronically exposed to iAs in drinking water [115], using a whole genome tiling-path array comparative genomic hybridization (CGH) platform [139]. It was detected a surprisingly low frequency of DNA gains at chromosome arm 3q in lung SqCCs from arsenic-exposed individuals (Figure 3), which is remarkable, since DNA gains at 3q are the most widely reported alteration associated with lung SqCC tumors and cell lines [140, 141].

It was also identified specific DNA gains and losses associated with lung SqCC from never smokers exposed to iAs. For example, a specific and frequent DNA gain at 19q13.33 contains genes related to ssDNA breaks repair process (POLD1) and neoplastic processes (SPIB and NR1H2). Additionally, a widespread association of DNA copy number loss specific to iAs-exposed lung SqCC, concordant with previous findings showing that arsenic can induce multiple large deletions through the creation of ROS [142] was identified. Some of these deletions, mainly at 9q12, may be relevant to iAs carcinogenic mechanisms, since they have been described in other iAs-related types of cancer and involve genes from the forkhead box (Fox) gene family, which have been linked to tumorigenesis and cancer progression [143].

This recent information provides evidence of distinct CNAs associated with lung SqCC occurring in patients who had exposure to iAs in drinking water and suggests that alternative molecular pathways are activated in this disease subset.

10. Conclusion

Arsenic exposure through contaminated drinking water poses a major health concern for over 40 million people worldwide, where for some, arsenic levels are almost 10 times higher than recommended thresholds. In addition to causing a variety of health problems including vascular and neurological conditions, arsenic is an established carcinogen. The rate of cancer incidence and mortality in populations exposed to arsenic contaminated drinking water is alarming. These populations experience particularly exacerbated rates of cancer in organs where arsenic is most concentrated or is excreted, including lung, bladder, and skin cancers. The mechanisms of arsenic-induced carcinogenesis are slowly being elucidated through the study of the precise DNA damaging and cytotoxic properties related to the biotransformation, metabolism, and excretion of arsenic. Discovery of particular genomic and epigenomic lesions induced by this metalloid should encourage a comprehensive approach to elucidate how arsenic can induce different types of cancer. Despite histology similarity, the possibility of iAs-induced cases biologically distinct entities, compared to those induced by other environmental carcinogens, must be considered. Knowledge related to these processes may lead to specific treatment strategies targeting arsenic-induced disorders and malignancies.

Acknowledgments

This work was supported by funds from the Canadian Institutes for Health Research (CIHR; MOP 86731, MOP 94867) and a CIHR CGS scholarship to E. A. Vucic. V. D. Martinez was supported by INNOVA-Chile grant 07CN13-PBT-48.

References

[1] C. Klein, *Mineral Science*, Wiley, New York, NY, USA, 2002.

[2] ATSDR, A.f.T.S.a.D.R., *Toxicological Profile for Arsenic, P.H.S.*, U.S. Department of Health and Human Services, 2007.

[3] J. M. Azcue, A. Mudroch, F. Rosa, G. E. M. Hall, T. A. Jackson, and T. Reynoldson, "Trace elements in water, sediments, porewater, and biota polluted by tailings from an abandoned gold mine in British Columbia, Canada," *Journal of Geochemical Exploration*, vol. 52, no. 1-2, pp. 25–34, 1995.

[4] P. L. Smedley and D. G. Kinniburgh, "A review of the source, behaviour and distribution of arsenic in natural waters," *Applied Geochemistry*, vol. 17, no. 5, pp. 517–568, 2002.

[5] M. Vahter and G. Concha, "Role of metabolism in arsenic toxicity," *Pharmacology and Toxicology*, vol. 89, no. 1, pp. 1–5, 2001.

[6] C. J. Chen, Y. M. Hsueh, M. S. Lai et al., "Increased prevalence of hypertension and long-term arsenic exposure," *Hypertension*, vol. 25, no. 1, pp. 53–60, 1995.

[7] R. R. Engel, C. Hopenhayn-Rich, O. Receveur, and A. H. Smith, "Vascular effects of chronic arsenic exposure: a review," *Epidemiologic Reviews*, vol. 16, no. 2, pp. 184–209, 1994.

[8] T. Kadono, T. Inaoka, N. Murayama et al., "Skin manifestations of arsenicosis in two villages in Bangladesh," *International Journal of Dermatology*, vol. 41, no. 12, pp. 841–846, 2002.

[9] W. P. Tseng, H. M. Chu, S. W. How, J. M. Fong, C. S. Lin, and S. Yeh, "Prevalence of skin cancer in an endemic area of chronic arsenicism in Taiwan," *Journal of the National Cancer Institute*, vol. 40, no. 3, pp. 453–463, 1968.

[10] C. J. Chen and C. J. Wang, "Ecological correlation between arsenic level in well water and age-adjusted mortality from malignant neoplasms," *Cancer Research*, vol. 50, no. 17, pp. 5470–5474, 1990.

[11] A. H. Smith, M. Goycolea, R. Haque, and M. L. Biggs, "Marked increase in bladder and lung cancer mortality in a region of northern chile due to arsenic in drinking water," *American Journal of Epidemiology*, vol. 147, no. 7, pp. 660–669, 1998.

[12] C. Hopenhayn-Rich, M. L. Biggs, A. Fuchs et al., "Bladder cancer mortality associated with arsenic in drinking water in Argentina," *Epidemiology*, vol. 7, no. 2, pp. 117–124, 1996.

[13] K. H. Morales, L. Ryan, T. L. Kuo, M. M. Wu, and C. J. Chen, "Risk of internal cancers from arsenic in drinking water," *Environmental Health Perspectives*, vol. 108, no. 7, pp. 655–661, 2000.

[14] C. Ferreccio, C. González, V. Milosavjlevic, G. Marshall, A. M. Sancha, and A. H. Smith, "Lung cancer and arsenic concentrations in drinking water in Chile," *Epidemiology*, vol. 11, no. 6, pp. 673–679, 2000.

[15] G. Marshall, C. Ferreccio, Y. Yuan et al., "Fifty-year study of lung and bladder cancer mortality in Chile related to arsenic in drinking water," *Journal of the National Cancer Institute*, vol. 99, no. 12, pp. 920–928, 2007.

[16] A. H. Smith, C. Hopenhayn-Rich, M. N. Bates et al., "Cancer risks from arsenic in drinking water," *Environmental Health Perspectives*, vol. 97, pp. 259–267, 1992.

[17] World health Organization (WHO), *Guidelines for Drinking-Water Quality*, 1993.

[18] U. S. Environmental Protection Agency, *Technical Fact Sheet: Final Rule for Arsenic in Drinking Water*, 2001.

[19] D. K. Nordstrom, "Worldwide occurrences of arsenic in ground water," *Science*, vol. 296, no. 5576, pp. 2143–2145, 2002.

[20] L. R. Croal, J. A. Gralnick, D. Malasarn, and D. K. Newman, "The genetics of geochemistry," *Annual Review of Genetics*, vol. 38, pp. 175–202, 2004.

[21] A. Basu, J. Mahata, S. Gupta, and A. K. Giri, "Genetic toxicology of a paradoxical human carcinogen, arsenic: A review," *Mutation Research*, vol. 488, no. 2, pp. 171–194, 2001.

[22] B. British Geological Survey, "British Geological Survey. Groundwater Studies for Arsenic Contamination in Bangladesh," Final Report, Mott MacDonald Ltd., 1999.

[23] R. Nickson, J. McArthur, W. Burgess, K. Matin Ahmed, P. Ravenscroft, and M. Rahman, "Arsenic poisoning of Bangladesh groundwater," *Nature*, vol. 395, no. 6700, p. 338, 1998.

[24] M. N. Mead, "Arsenic: in search of an antidote to a global poison," *Environmental Health Perspectives*, vol. 113, no. 6, pp. A378–A386, 2005.

[25] W. P. Tseng, "Effects and dose response relationships of skin cancer and blackfoot disease with arsenic," *Environmental Health Perspectives*, vol. 19, pp. 109–119, 1977.

[26] C.-J. Chen, M.-M. Wu, S.-S. Lee, J.-D. Wang, S.-H. Cheng, and H.-Y. Wu, "Atherogenicity and carcinogenicity of high-arsenic artesian well water. Multiple risk factors and related malignant neoplasms of blackfoot disease," *Arteriosclerosis*, vol. 8, no. 5, pp. 452–460, 1988.

[27] R. Blackwell, "Estimation total arsenic ingested by residents in Ihe endemic blackfoot area," *Journal of the Formosan Medical Association*, vol. 60, pp. 1143–1144, 1961.

[28] G. B. Freeman, R. A. Schoof, M. V. Ruby et al., "Bioavailability of arsenic in soil and house dust impacted by smelter activities following oral administration in cynomolgus monkeys," *Fundamental and Applied Toxicology*, vol. 28, no. 2, pp. 215–222, 1995.

[29] C. Pomroy, S. M. Charbonneau, R. S. McCullough, and G. K. H. Tam, "Human retention studies with 74As," *Toxicology and Applied Pharmacology*, vol. 53, no. 3, pp. 550–556, 1980.

[30] M. Vahter and H. Norin, "Metabolism of 74As-labeled trivalent and pentavalent inorganic arsenic in mice," *Environmental Research*, vol. 21, no. 2, pp. 446–457, 1980.

[31] World health Organization (WHO), *Air Quality Guidelines*, W.R. Publications, WHO Regional Office for Europe, Copenhagen, Denmark, 2nd edition, 2000.

[32] E. Marafante, M. Vahter, and J. Envall, "The role of the methylation in the detoxication of arsenate in the rabbit," *Chemico-Biological Interactions*, vol. 56, no. 2-3, pp. 225–238, 1985.

[33] M. Vahter and E. Marafante, "Reduction and binding of arsenate in marmoset monkeys," *Archives of Toxicology*, vol. 57, no. 2, pp. 119–124, 1985.

[34] M. Vahter, "Methylation of inorganic arsenic in different mammalian species and population groups," *Science Progress*, vol. 82, part 1, pp. 69–88, 1999.

[35] M. Vahter, "Genetic polymorphism in the biotransformation of inorganic arsenic and its role in toxicity," *Toxicology Letters*, vol. 112-113, pp. 209–217, 2000.

[36] D. J. Thompson, "A chemical hypothesis for arsenic methylation in mammals," *Chemico-Biological Interactions*, vol. 88, no. 2-3, pp. 89–114, 1993.

[37] S. Lin, Q. Shi, F. Brent Nix et al., "A novel S-adenosyl-L-methionine:arsenic(III) methyltransferase from rat liver cytosol," *Journal of Biological Chemistry*, vol. 277, no. 13, pp. 10795–10803, 2002.

[38] B. K. Mandal, Y. Ogra, and K. T. Suzuki, "Identification of dimethylarsinous and monomethylarsonous acids in human urine of the arsenic-affected areas in West Bengal, India," *Chemical Research in Toxicology*, vol. 14, no. 4, pp. 371–378, 2001.

[39] H. V. Aposhian, E. S. Gurzau, X. C. Le et al., "Occurrence of monomethylarsonous acid in urine of humans exposed to inorganic arsenic," *Chemical Research in Toxicology*, vol. 13, no. 8, pp. 693–697, 2000.

[40] M. M. Meza, M. J. Kopplin, J. L. Burgess, and A. J. Gandolfi, "Arsenic drinking water exposure and urinary excretion among adults in the Yaqui Valley, Sonora, Mexico," *Environmental Research*, vol. 96, no. 2, pp. 119–126, 2004.

[41] T. Hayakawa, Y. Kobayashi, X. Cui, and S. Hirano, "A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19," *Archives of Toxicology*, vol. 79, no. 4, pp. 183–191, 2005.

[42] C. Kojima, D. C. Ramirez, E. J. Tokar et al., "Requirement of arsenic biomethylation for oxidative DNA damage," *Journal of the National Cancer Institute*, vol. 101, no. 24, pp. 1670–1681, 2009.

[43] M. I. Ekstrand, M. Falkenberg, A. Rantanen et al., "Mitochondrial transcription factor A regulates mtDNA copy number in mammals," *Human Molecular Genetics*, vol. 13, no. 9, pp. 935–944, 2004.

[44] T. Kanki, K. Ohgaki, M. Gaspari et al., "Architectural role of mitochondrial transcription factor a in maintenance of human mitochondrial DNA," *Molecular and Cellular Biology*, vol. 24, no. 22, pp. 9823–9834, 2004.

[45] K. P. Singh, R. Kumari, J. Treas, and J. W. Dumond, "Chronic exposure to arsenic causes increased cell survival, DNA damage, and increased expression of mitochondrial transcription factor A (mtTFA) in human prostate epithelial cells," *Chemical Research in Toxicology*, vol. 24, no. 3, pp. 340–349, 2011.

[46] T. G. Rossman, "Mechanism of arsenic carcinogenesis: an integrated approach," *Mutation Research*, vol. 533, no. 1-2, pp. 37–65, 2003.

[47] J. C. Barrett, P. W. Lamb, T. C. Wang, and T. C. Lee, "Mechanism of arsenic-induced cell transformation," *Biological Trace Element Research*, vol. 21, pp. 421–429, 1989.

[48] P. P. Simeonova and M. I. Luster, "Mechanisms of arsenic carcinogenicity: genetic or epigenetic mechanisms?" *Journal of Environmental Pathology, Toxicology and Oncology*, vol. 19, no. 3, pp. 281–286, 2000.

[49] M. Styblo, L. M. Del Razo, L. Vega et al., "Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells," *Archives of Toxicology*, vol. 74, no. 6, pp. 289–299, 2000.

[50] J. S. Petrick, F. Ayala-Fierro, W. R. Cullen, D. E. Carter, and H. Vasken Aposhian, "Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes," *Toxicology and Applied Pharmacology*, vol. 163, no. 2, pp. 203–207, 2000.

[51] M. J. Mass, A. Tennant, B. C. Roop et al., "Methylated trivalent arsenic species are genotoxic," *Chemical Research in Toxicology*, vol. 14, no. 4, pp. 355–361, 2001.

[52] S. Ahmad, K. T. Kitchin, and W. R. Cullen, "Arsenic species that cause release of iron from ferritin and generation of activated oxygen," *Archives of Biochemistry and Biophysics*, vol. 382, no. 2, pp. 195–202, 2000.

[53] K. Yamanaka, M. Mizoi, M. Tachikawa, A. Hasegawa, M. Hoshino, and S. Okada, "Oxidative DNA damage following exposure to dimethylarsinous iodide: the formation of cis-thymine glycol," *Toxicology Letters*, vol. 143, no. 2, pp. 145–153, 2003.

[54] K. Yamanaka, M. Mizoi, K. Kato, A. Hasegawa, M. Nakano, and S. Okada, "Oral administration of dimethylarsinic acid, a main metabolite of inorganic arsenic, in mice promotes skin tumorigenesis initiated by dimethylbenz(a)anthracene with or without ultraviolet B as a promoter," *Biological and Pharmaceutical Bulletin*, vol. 24, no. 5, pp. 510–514, 2001.

[55] K. Yamanaka, K. Katsumata, K. Ikuma, A. Hasegawa, M. Nakano, and S. Okada, "The role of orally administered dimethylarsinic acid, a main metabolite of inorganic arsenics, in the promotion and progression of UVB-induced skin tumorigenesis in hairless mice," *Cancer Letters*, vol. 152, no. 1, pp. 79–85, 2000.

[56] M. Wei, H. Wanibuchi, S. Yamamoto, W. Li, and S. Fukushima, "Urinary bladder carcinogenicity of dimethylarsinic acid in male F344 rats," *Carcinogenesis*, vol. 20, no. 9, pp. 1873–1876, 1999.

[57] H. Wanibuchi, S. Yamamoto, H. Chen et al., "Promoting effects of dimethylarsinic acid on N-butyl-N-(4-hydroxybutyl)nitrosamine-induced urinary bladder carcinogenesis in rats," *Carcinogenesis*, vol. 17, no. 11, pp. 2435–2439, 1996.

[58] S. Yamamoto, Y. Konishi, T. Matsuda et al., "Cancer induction by an organic arsenic compound, dimethylarsinic acid (cacodylic acid), in F344/DuCrj rats after pretreatment with five carcinogens," *Cancer Research*, vol. 55, no. 6, pp. 1271–1276, 1995.

[59] M. Kessel, S. X. Liu, A. Xu, R. Santella, and T. K. Hei, "Arsenic induces oxidative DNA damage in mammalian cells," *Molecular and Cellular Biochemistry*, vol. 234-235, pp. 301–308, 2002.

[60] K. T. Kitchin and S. Ahmad, "Oxidative stress as a possible mode of action for arsenic carcinogenesis," *Toxicology Letters*, vol. 137, no. 1-2, pp. 3–13, 2003.

[61] H. Shi, X. Shi, and K. J. Liu, "Oxidative mechanism of arsenic toxicity and carcinogenesis," *Molecular and Cellular Biochemistry*, vol. 255, no. 1-2, pp. 67–78, 2004.

[62] R. C. Lantz and A. M. Hays, "Role of oxidative stress in arsenic-induced toxicity," *Drug Metabolism Reviews*, vol. 38, no. 4, pp. 791–804, 2006.

[63] M. Valko, C. J. Rhodes, J. Moncol, M. Izakovic, and M. Mazur, "Free radicals, metals and antioxidants in oxidative stress-induced cancer," *Chemico-Biological Interactions*, vol. 160, no. 1, pp. 1–40, 2006.

[64] A. De Vizcaya-Ruiz, O. Barbier, R. Ruiz-Ramos, and M. E. Cebrian, "Biomarkers of oxidative stress and damage in human populations exposed to arsenic," *Mutation Research*, vol. 674, no. 1-2, pp. 85–92, 2009.

[65] K. T. Kitchin and R. Conolly, "Arsenic-induced carcinogenesissoxidative stress as a possible mode of action and future research needs for more biologically based risk assessment," *Chemical Research in Toxicology*, vol. 23, no. 2, pp. 327–335, 2010.

[66] A. Barchowsky, L. R. Klei, E. J. Dudek, H. M. Swartz, and P. E. James, "Stimulation of reactive oxygen, but not reactive nitrogen species, in vascular endothelial cells exposed to low

levels of arsenite," *Free Radical Biology and Medicine*, vol. 27, no. 11-12, pp. 1405–1412, 1999.

[67] F. Liu and K. Y. Jan, "DNA damage in arsenite- and cadmium-treated bovine aortic en-do-the-lial cells," *Free Radical Biology and Medicine*, vol. 28, no. 1, pp. 55–63, 2000.

[68] S. J. S. Flora, "Arsenic-induced oxidative stress and its reversibility following combined administration of N-acetylcysteine and meso 2,3-dimercaptosuccinic acid in rats," *Clinical and Experimental Pharmacology and Physiology*, vol. 26, no. 11, pp. 865–869, 1999.

[69] S. X. Liu, M. Athar, I. Lippai, C. Waldren, and T. K. Hei, "Induction of oxyradicals by arsenic: implication for mechanism of genotoxicity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 4, pp. 1643–1648, 2001.

[70] K. Yamanaka and S. Okada, "Induction of lung-specific DNA damage by metabolically methylated arsenics via the production of free radicals," *Environmental Health Perspectives*, vol. 102, no. 3, pp. 37–40, 1994.

[71] E. M. Kenyon, M. F. Hughes, B. M. Adair et al., "Tissue distribution and urinary excretion of inorganic arsenic and its methylated metabolites in C57BL6 mice following subchronic exposure to arsenate in drinking water," *Toxicology and Applied Pharmacology*, vol. 232, no. 3, pp. 448–455, 2008.

[72] S. Ahmad, K. T. Kitchin, and W. R. Cullen, "Plasmid DNA damage caused by methylated arsenicals, ascorbic acid and human liver ferritin," *Toxicology Letters*, vol. 133, no. 1, pp. 47–57, 2002.

[73] M. A. Partridge, S. X. L. Huang, E. Hernandez-Rosa, M. M. Davidson, and T. K. Hei, "Arsenic induced mitochondrial DNA damage and altered mitochondrial oxidative function: implications for genotoxic mechanisms in mammalian cells," *Cancer Research*, vol. 67, no. 11, pp. 5239–5247, 2007.

[74] S. X. Liu, M. M. Davidson, X. Tang et al., "Mitochondrial damage mediates genotoxicity of arsenic in mammalian cells," *Cancer Research*, vol. 65, no. 8, pp. 3236–3242, 2005.

[75] S. H. Woo, I. C. Park, M. J. Park et al., "Arsenic trioxide induces apoptosis through a reactive oxygen species-dependent pathway and loss of mitochondrial membrane potential in HeLa cells," *International Journal of Oncology*, vol. 21, no. 1, pp. 57–63, 2002.

[76] S. J. Flora, "Arsenic-induced oxidative stress and its reversibility," *Free Radical Biology & Medicine*, vol. 51, no. 2, pp. 257–281, 2011.

[77] J. H. Li and T. G. Rossman, "Comutagenesis of sodium arsenite with ultraviolet radiation in Chinese hamster V79 cells," *Biology of Metals*, vol. 4, no. 4, pp. 197–200, 1991.

[78] T. K. Hei, S. U. X. Liu, and C. Waldren, "Mutagenicity of arsenic in mammalian cells: role of reactive oxygen species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 14, pp. 8103–8107, 1998.

[79] K. P. Singh and J. W. DuMond, "Genetic and epigenetic changes induced by chronic low dose exposure to arsenic of mouse testicular Leydig cells," *International Journal of Oncology*, vol. 30, no. 1, pp. 253–260, 2007.

[80] A. D. Kligerman, C. L. Doerr, A. H. Tennant et al., "Methylated trivalent arsenicals as candidate ultimate genotoxic forms of arsenic: induction of chromosomal mutations but not gene mutations," *Environmental and Molecular Mutagenesis*, vol. 42, no. 3, pp. 192–205, 2003.

[81] K. T. Kitchin and K. Wallace, "Evidence against the nuclear in situ binding of arsenicals-oxidative stress theory of arsenic carcinogenesis," *Toxicology and Applied Pharmacology*, vol. 232, no. 2, pp. 252–257, 2008.

[82] B. Halliwell, "Oxidative stress and cancer: have we moved forward?" *Biochemical Journal*, vol. 401, no. 1, pp. 1–11, 2007.

[83] K. W. Caldecott, "DNA single-strand breaks and neurodegeneration," *DNA Repair*, vol. 3, no. 8-9, pp. 875–882, 2004.

[84] A. D. Kligerman, S. I. Malik, and J. A. Campbell, "Cytogenetic insights into DNA damage and repair of lesions induced by a monomethylated trivalent arsenical.," *Mutation Research*, vol. 695, no. 1-2, pp. 2–8, 2010.

[85] S. A. Mourón, C. A. Grillo, F. N. Dulout, and C. D. Golijow, "Induction of DNA strand breaks, DNA-protein crosslinks and sister chromatid exchanges by arsenite in a human lung cell line," *Toxicology in Vitro*, vol. 20, no. 3, pp. 279–285, 2006.

[86] S. Ying, K. Myers, S. Bottomley, T. Helleday, and H. E. Bryant, "BRCA2-dependent homologous recombination is required for repair of Arsenite-induced replication lesions in mammalian cells," *Nucleic Acids Research*, vol. 37, no. 15, pp. 5105–5113, 2009.

[87] T. Helleday, J. Lo, D. C. van Gent, and B. P. Engelward, "DNA double-strand break repair: from mechanistic understanding to cancer treatment," *DNA Repair*, vol. 6, no. 7, pp. 923–935, 2007.

[88] X. J. Qin, L. G. Hudson, W. Liu, G. S. Timmins, and K. J. Liu, "Low concentration of arsenite exacerbates UVR-induced DNA strand breaks by inhibiting PARP-1 activity," *Toxicology and Applied Pharmacology*, vol. 232, no. 1, pp. 41–50, 2008.

[89] A. D. Kligerman and A. H. Tennant, "Insights into the carcinogenic mode of action of arsenic," *Toxicology and Applied Pharmacology*, vol. 222, no. 3, pp. 281–288, 2007.

[90] T. C. Lee, N. Tanaka, P. W. Lamb, T. M. Gilmer, and J. C. Barrett, "Induction of gene amplification by arsenic," *Science*, vol. 241, no. 4861, pp. 79–81, 1988.

[91] F. W. Alt, R. E. Kellems, J. R. Bertino, and R. T. Schimke, "Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells," *Journal of Biological Chemistry*, vol. 253, no. 5, pp. 1357–1370, 1978.

[92] L. E. Moore, A. H. Smith, C. Eng et al., "Arsenic-related chromosomal alterations in bladder cancer," *Journal of the National Cancer Institute*, vol. 94, no. 22, pp. 1688–1696, 2002.

[93] L. I. Hsu, A. W. Chiu, Y. S. Pu et al., "Comparative genomic hybridization study of arsenic-exposed and non-arsenic-exposed urinary transitional cell carcinoma," *Toxicology and Applied Pharmacology*, vol. 227, no. 2, pp. 229–238, 2008.

[94] M. E. Gonsebatt, L. Vega, A. M. Salazar et al., "Cytogenetic effects in human exposure to arsenic," *Mutation Research*, vol. 386, no. 3, pp. 219–228, 1997.

[95] F. N. Dulout, C. A. Grillo, A. I. Seoane et al., "Chromosomal aberrations in peripheral blood lymphocytes from native Andean women and children from Northwestern Argentina exposed to arsenic in drinking water," *Mutation Research*, vol. 370, no. 3-4, pp. 151–158, 1996.

[96] Y. An, K. Kato, M. Nakano, H. Otsu, S. Okada, and K. Yamanaka, "Specific induction of oxidative stress in terminal bronchiolar Clara cells during dimethylarsenic-induced lung tumor promoting process in mice," *Cancer Letters*, vol. 230, no. 1, pp. 57–64, 2005.

[97] Y. An, Z. Gao, Z. Wang et al., "Immunohistochemical analysis of oxidative DNA damage in arsenic-related human skin

samples from arsenic-contaminated area of China," *Cancer Letters*, vol. 214, no. 1, pp. 11–18, 2004.

[98] M. Matsui, C. Nishigori, S. Toyokuni et al., "The role of oxidative DNA damage in human arsenic carcinogenesis: detection of 8-hydroxy-2′-deoxyguanosine in arsenic-related Bowen's disease," *Journal of Investigative Dermatology*, vol. 113, no. 1, pp. 26–31, 1999.

[99] H. Wanibuchi, T. Hori, V. Meenakshi et al., "Promotion of rat hepatocarcinogenesis by dimethylarsinic acid: association with elevated ornithine decarboxylase activity and formation of 8-hydroxydeoxyguanosine in the liver," *Japanese Journal of Cancer Research*, vol. 88, no. 12, pp. 1149–1154, 1997.

[100] R. Ruiz-Ramos, L. Lopez-Carrillo, A. D. Rios-Perez, A. De Vizcaya-Ruíz, and M. E. Cebrian, "Sodium arsenite induces ROS generation, DNA oxidative damage, HO-1 and c-Myc proteins, NF-κB activation and cell proliferation in human breast cancer MCF-7 cells," *Mutation Research*, vol. 674, no. 1-2, pp. 109–115, 2009.

[101] A. Hartwig, H. Blessing, T. Schwerdtle, and I. Walter, "Modulation of DNA repair processes by arsenic and selenium compounds," *Toxicology*, vol. 193, no. 1-2, pp. 161–167, 2003.

[102] A. E. O. Fisher, H. Hochegger, S. Takeda, and K. W. Caldecott, "Poly(ADP-ribose) polymerase 1 accelerates single-strand break repair in concert with poly(ADP-ribose) glycohydrolase," *Molecular and Cellular Biology*, vol. 27, no. 15, pp. 5597–5605, 2007.

[103] K. Yamanaka, H. Hayashi, K. Kalo, A. Hasegawa, and S. Okada, "Involvement of pref-er-ential formation of apurinic/apyrimidinic sites in dimethylarsenic-induced DNA strand breaks and DNA-protein crosslinks in cultured alveolar epithelial cells," *Biochemical and Biophysical Research Communications*, vol. 207, no. 1, pp. 244–249, 1995.

[104] K. Yamanaka, H. Hayashi, M. Tachikawa et al., "Metabolic methylation is a possible genotoxicity-enhancing process of inorganic arsenics," *Mutation Research*, vol. 394, no. 1–3, pp. 95–101, 1997.

[105] J. de Boer and J. H. J. Hoeijmakers, "Nucleotide excision repair and human syndromes," *Carcinogenesis*, vol. 21, no. 3, pp. 453–460, 2000.

[106] T. Schwerdtle, I. Walter, and A. Hartwig, "Arsenite and its biomethylated metabolites interfere with the formation and repair of stable BPDE-induced DNA adducts in human cells and impair XPAzf and Fpg," *DNA Repair*, vol. 2, no. 12, pp. 1449–1463, 2003.

[107] M. J. Osmond, B. A. Kunz, and E. T. Snow, "Age and exposure to arsenic alter base excision repair transcript levels in mice," *Mutagenesis*, vol. 25, no. 5, pp. 517–522, 2010.

[108] P. Sykora and E. T. Snow, "Modulation of DNA polymerase beta-dependent base excision repair in cultured human cells after low dose exposure to arsenite," *Toxicology and Applied Pharmacology*, vol. 228, no. 3, pp. 385–394, 2008.

[109] E. T. Snow, Y. Hu, C. B. Klein, K. L. McCluskey, M. Schuliga, and P. Sykora, "Regulation of redox and DNA repair genes by arsenic: low dose protection against oxidative stress?" in *Arsenic Exposure and Health Effects V*, W. R. Chappel, C. O. Abernathy, R. L. Calderon, and D. J. Thomas, Eds., pp. 305–319, Elsevier Science, San Diego, Calif, USA, 2003.

[110] E. T. Snow, Y. Hu, C. C. Yan, and S. Chouchane, "Modulation of DNA repair and glutathione levels in human keratinocytes by micromolar arsenite," in *Proceedings of the 3rd International Conference on Arsenic Exposure and Health Effects*, W. R. Chappel, C. O. Abernathy, and R. L. Calderon, Eds., pp. 243–251, Elsevier Science, Oxford, UK, 1999.

[111] T. C. Zhang, M. T. Schmitt, and J. L. Mumford, "Effects of arsenic on telomerase and telomeres in relation to cell proliferation and apoptosis in human keratinocytes and leukemia cells in vitro," *Carcinogenesis*, vol. 24, no. 11, pp. 1811–1817, 2003.

[112] D. C. Cabelof, J. J. Raffoul, S. Yanamadala, Z. Guo, and A. R. Heydari, "Induction of DNA polymerase β-dependent base excision repair in response to oxidative stress in vivo," *Carcinogenesis*, vol. 23, no. 9, pp. 1419–1425, 2002.

[113] K. H. Chen, D. K. Srivastava, R. K. Singhal, S. Jacob, A. E. Ahmed, and S. H. Wilson, "Modulation of base excision repair by low density lipoprotein, oxidized low density lipoprotein and antioxidants in mouse monocytes," *Carcinogenesis*, vol. 21, no. 5, pp. 1017–1022, 2000.

[114] K. H. Chen, F. M. Yakes, D. K. Srivastava et al., "Up-regulation of base excision repair correlates with enhanced protection against a DNA damaging agent in mouse cell lines," *Nucleic Acids Research*, vol. 26, no. 8, pp. 2001–2007, 1998.

[115] V. D. Martinez, T. P. H. Buys, M. Adonis et al., "Arsenic-related DNA copy-number alterations in lung squamous cell carcinomas," *British Journal of Cancer*, vol. 103, no. 8, pp. 1277–1283, 2010.

[116] C. Nishida, P. Reinhard, and S. Linn, "DNA repair synthesis in human fibroblasts requires DNA plymerase δ," *Journal of Biological Chemistry*, vol. 263, no. 1, pp. 501–510, 1988.

[117] J. L. Parsons, B. D. Preston, T. R. O'Connor, and G. L. Dianov, "DNA polymerase δ-dependent repair of DNA single strand breaks containing 3′-end proximal lesions," *Nucleic Acids Research*, vol. 35, no. 4, pp. 1054–1063, 2007.

[118] R. E. Goldsby, L. E. Hays, X. Chen et al., "High incidence of epithelial cancers in mice deficient for DNA polymerase δ proofreading," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 24, pp. 15560–15565, 2002.

[119] R. N. Venkatesan, P. M. Treuting, E. D. Fuller et al., "Mutation at the polymerase active site of mouse DNA polymerase δ increases genomic instability and accelerates tumorigenesis," *Molecular and Cellular Biology*, vol. 27, no. 21, pp. 7669–7682, 2007.

[120] H. K. Hamadeh, K. J. Trouba, R. P. Amin, C. A. Afshari, and D. Germolec, "Coordination of altered DNA repair and damage pathways in arsenite-exposed keratinocytes," *Toxicological Sciences*, vol. 69, no. 2, pp. 306–316, 2002.

[121] A. S. Andrew, M. R. Karagas, and J. W. Hamilton, "Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water," *International Journal of Cancer*, vol. 104, no. 3, pp. 263–268, 2003.

[122] A. S. Andrew, J. L. Burgess, M. M. Meza et al., "Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic," *Environmental Health Perspectives*, vol. 114, no. 8, pp. 1193–1198, 2006.

[123] J. Mo, Y. Xia, Z. Ning, T. J. Wade, and J. L. Mumford, "Elevated ERCC1 gene expression in blood cells associated with exposure to arsenic from drinking water in inner Mongolia," *Anticancer Research*, vol. 29, no. 8, pp. 3253–3259, 2009.

[124] A. Memisoglu and L. Samson, "Base excision repair in yeast and mammals," *Mutation Research*, vol. 451, no. 1-2, pp. 39–51, 2000.

[125] J. Mo, Y. Xia, T. J. Wade et al., "Chronic arsenic exposure and oxidative stress: OGG1 expression and arsenic exposure,

nail selenium, and skin hyperkeratosis in inner Mongolia," *Environmental Health Perspectives*, vol. 114, no. 6, pp. 835–841, 2006.

[126] T. Lindahl and R. D. Wood, "Quality control by DNA repair," *Science*, vol. 286, no. 5446, pp. 1897–1905, 1999.

[127] S. Lynn, H. T. Lai, J. R. Gurr, and K. Y. Jan, "Arsenite retards DNA break rejoining by inhibiting DNA ligation," *Mutagenesis*, vol. 12, no. 5, pp. 353–358, 1997.

[128] Y. Hu, L. Su, and E. T. Snow, "Arsenic toxicity is enzyme specific and its affects on ligation are not caused by the direct inhibition of DNA repair enzymes," *Mutation Research*, vol. 408, no. 3, pp. 203–218, 1998.

[129] J. Liu, M. B. Kadiiska, Y. Liu, T. Lu, W. Qu, and M. P. Waalkes, "Stress-related gene expression in mice treated with inorganic arsenicals," *Toxicological Sciences*, vol. 61, no. 2, pp. 314–320, 2001.

[130] S. Shall and G. De Murcia, "Poly(ADP-ribose) polymerase-1: what have we learned from the deficient mouse model?" *Mutation Research*, vol. 460, no. 1, pp. 1–15, 2000.

[131] A. Bürkle, J. Diefenbach, C. Brabeck, and S. Beneke, "Ageing and PARP," *Pharmacological Research*, vol. 52, no. 1, pp. 93–99, 2005.

[132] V. Schreiber, F. Dantzer, J. C. Amé, and G. De Murcia, "Poly(ADP-ribose): novel functions for an old molecule," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 7, pp. 517–528, 2006.

[133] A. Poonepalli, L. Balakrishnan, A. K. Khaw et al., "Lack of poly(ADP-ribose) polymerase-1 gene product enhances cellular sensitivity to arsenite," *Cancer Research*, vol. 65, no. 23, pp. 10977–10983, 2005.

[134] F. Le Page, V. Schreiber, C. Dhérin, G. De Murcia, and S. Boiteux, "Poly(ADP-ribose) polymerase-1 (PARP-1) is required in murine cell lines for base excision repair of oxidative DNA damage in the absence of DNA polymerase β," *Journal of Biological Chemistry*, vol. 278, no. 20, pp. 18471–18477, 2003.

[135] S. Petersen, G. Saretzki, and T. von Zglinicki, "Preferential accumulation of single-stranded regions in telomeres of human fibroblasts," *Experimental Cell Research*, vol. 239, no. 1, pp. 152–160, 1998.

[136] F. Dantzer, G. De La Rubia, J. Ménissier-De Murcia, Z. Hostomsky, G. De Murcia, and V. Schreiber, "Base excision repair is impaired in mammalian cells lacking poly(ADP-ribose) polymerase-1," *Biochemistry*, vol. 39, no. 25, pp. 7559–7569, 2000.

[137] H. Gao, D. L. Coyle, M. L. Meyer-Ficca et al., "Altered poly(ADP-ribose) metabolism impairs cellular responses to genotoxic stress in a hypomorphic mutant of poly(ADP-ribose) glycohydrolase," *Experimental Cell Research*, vol. 313, no. 5, pp. 984–996, 2007.

[138] S. Sun, J. H. Schiller, and A. F. Gazdar, "Lung cancer in never smokers—a different disease," *Nature Reviews Cancer*, vol. 7, no. 10, pp. 778–790, 2007.

[139] S. K. Watson, R. J. deLeeuw, D. E. Horsman, J. A. Squire, and W. L. Lam, "Cytogenetically balanced translocations are associated with focal copy number alterations," *Human Genetics*, vol. 120, no. 6, pp. 795–805, 2007.

[140] G. Tonon, K. K. Wong, G. Maulik et al., "High-resolution genomic profiles of human lung cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9625–9630, 2005.

[141] C. Garnis, W. W. Lockwood, E. Vucic et al., "High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH," *International Journal of Cancer*, vol. 118, no. 6, pp. 1556–1564, 2006.

[142] G. Concha, G. Vogler, D. Lezcano, B. Nermell, and M. Vahter, "Exposure to inorganic arsenic metabolites during early human development," *Toxicological Sciences*, vol. 44, no. 2, pp. 185–190, 1998.

[143] I. M. Kim, T. Ackerson, S. Ramakrishna et al., "The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer," *Cancer Research*, vol. 66, no. 4, pp. 2153–2161, 2006.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 691735, 9 pages
doi:10.4061/2011/691735

Review Article

Structure and Function of the Small MutS-Related Domain

Kenji Fukui[1] and Seiki Kuramitsu[1,2]

[1] *RIKEN SPring-8 Center, Harima Institute, 1-1-1, Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan*
[2] *Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1, Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan*

Correspondence should be addressed to Seiki Kuramitsu, kuramitu@bio.sci.osaka-u.ac.jp

Received 1 March 2011; Accepted 13 May 2011

Academic Editor: Bernardo Reina-San-Martin

MutS family proteins are widely distributed in almost all organisms from bacteria to human and play central roles in various DNA transactions such as DNA mismatch repair and recombinational events. The small MutS-related (Smr) domain was originally found in the C-terminal domain of an antirecombination protein, MutS2, a member of the MutS family. MutS2 is thought to suppress homologous recombination by endonucleolytic resolution of early intermediates in the process. The endonuclease activity of MutS2 is derived from the Smr domain. Interestingly, sequences homologous to the Smr domain are abundant in a variety of proteins other than MutS2 and can be classified into 3 subfamilies. Recently, the tertiary structures and endonuclease activities of all 3 Smr subfamilies were reported. In this paper, we review the biochemical characteristics and structures of the Smr domains as well as cellular functions of the Smr-containing proteins.

1. Introduction

MutS is a key enzyme in DNA mismatch repair (MMR) that corrects mismatched bases produced during DNA replication and other biological processes [1, 2]. MutS recognizes mismatches and stimulates the downstream reactions in MMR [3–5]. MutS orthologues are conserved in almost all organisms including viruses, archaea, bacteria, and eukaryotes [6]. Accumulating genome information has revealed that proteins partly homologous to MutS also exist in all 3 domains of life [7–9]. Among those MutS paralogues, bacterial MutS2 has been relatively well characterized [10, 11]. Although disruption of *mutS2* in *Bacillus subtilis* and *Deinococcus radiodurans* did not affect the phenotype of either strain [12, 13], it was reported that knockout of *mutS2* in *Helicobacter pylori* caused an increase in the frequency of homologous recombination [10, 11]. This result suggests an inhibitory role for MutS2 in homologous recombination. Biochemical characterization demonstrated that, in *Helicobacter pylori* and *Thermus thermophilus*, MutS2 preferably binds to branched DNA structures, including Holliday junction and D-loop structure [10, 11, 14], the intermediates in homologous recombination. Finally, it was confirmed that *T. thermophilus* MutS2 endonucleolytically digests those branched DNA structures [14], indicating that MutS2 suppresses homologous recombination through the resolution of early intermediates in the process.

The endonuclease domain is located in the C-terminal region of MutS2, which is called the small MutS-related (Smr) domain [15]. While the Smr domain is not found in other MutS homologues, as initially pointed out by Moreira and Philippe [15], sequences homologous to the Smr domain are ubiquitous among a variety of proteins and are conserved in almost all organisms except in archaea (Table 1) [11, 16, 17]. Malik and Henikoff [17] predicted that the Smr domain has an endonuclease activity based on the domain-architecture analogy of MutS2 with *Sgmt*-MutS, a MutS homologue present in *Sarcophyton glaucum* mitochondria. *Sgmt*-MutS contains an HNH endonuclease domain in its C-terminal region [9]. Although the Smr domain has no sequence similarity with the HNH endonuclease domain, the endonuclease activity of Smr domain has been verified by an accumulating amount of experimental evidences [18–20]. In this paper, we review the recent reports about the functional and structural characterization of the Smr domains and discuss the cellular functions of Smr-containing proteins.

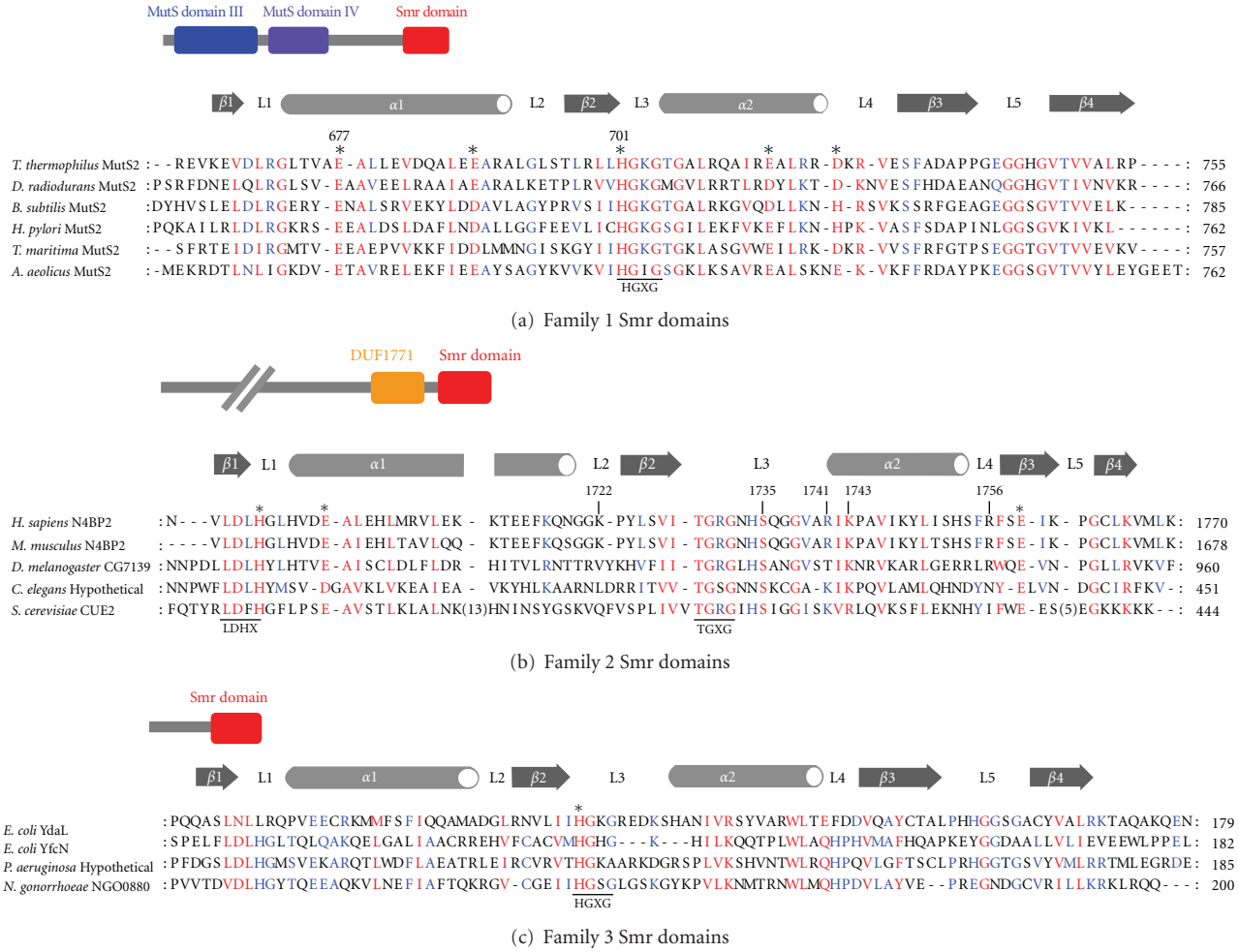
Table 1: Distribution of Smr domain-containing proteins.

Species	Proteins containing family 1 Smr domains	Proteins containing family 2 Smr domains	Proteins containing family 3 Smr domains
Thermus thermophilus	MutS2 (YP_144911)*[1]	—	—
Helicobacter pylori	MutS2 (ZP_03440043)	—	—
Bacillus subtilis	MutS2 (NP_390736)	—	—
Deinococcus radiodurans	MutS2 (NP_295699)	—	—
Aquifex aeolicus	MutS2 (NP_213851)	—	—
Thermotoga maritima	MutS2 (NP_229083)	—	—
Arabidopsis thaliana	MutS2 (NP_200220)	GUN1 (NP_849962) pTAC2 (NP_177623) SVR7 (Q8GWE0.2) At5G46580 (NP_199470) At1G79490 (NP_178067) At1G18900 (NP_973860) At1G74750 (NP_177613) At2G17033 (NP_849962)	—
Escherichia coli	—	—	YdaL (NP_415856)*[5] YfcN (AP_002931)
Neisseria gonorrhoeae	—	—	NGO0880 (YP_207992)
Pseudomonas aeruginosa	—	—	Hypothetical protein (AAG_05064)
Saccharomyces cerevisiae	—	CUE2 protein (EEU_05137) Ypl199cp (NP_015125)*[2]	—
Caenorhabditis elegans	—	Hypothetical protein (NP_498004)	Hypothetical protein (NP_494494) Hypothetical protein (NP_494390)
Drosophila melanogaster	—	CG7139, isoform A (NP_649378)*[3]	—
Mus musculus	—	N4BP2 (NP_001020088)*[4]	—
Homo sapiens	—	N4BP2 (NP_060647)	—

*[1] Numbers in parenthesis indicate accession numbers.
*[2] Ypl199cp also contains DUF1771.
*[3] CG7139 shows amino acid sequence similarity to mammalian N4BP2-like proteins. The highly conserved domain, DUF1771, is adjacent to the Smr domain in CG7139.
*[4] DUF1771 is adjacent to the Smr domains in mammalian N4BP2.
*[5] Organisms that possess a family 3 Smr domain do not have a family 1 Smr domain [11].

2. Smr Domains Can Be Categorized into 3 Subfamilies

Smr domains can be categorized roughly into 3 subfamilies on the basis of the domain architecture of the proteins in which the Smr domains are included (Figure 1). The first of these is the C-terminal domain of the MutS2 protein that is found in *Firmicutes*, *Deinococcus-Thermus*, *Bacteroidetes*, *Deltaproteobacteria*, and *Epsilonproteobacteria* phyla of bacteria and plants. Plant MutS2 is believed to be derived from the genomes of incorporated cyanobacteria [7]. In this paper, we refer to this type of Smr domain as a family 1 Smr domain (Figure 1(a)). The second subfamily consists of the C-terminal domains of proteins other than MutS2. For example, the C-terminal domain of the human NEDD4-binding protein 2 (N4BP2) has a significant amino acid sequence homology with the Smr domain of MutS2 [21]. We refer to this type as family 2 Smr domains (Figure 1(b)). The family 2 Smr domains are usually found in eukaryotes. The last group consist of the stand-alone type Smr domains, such as *E. coli* YdaL and *E. coli* YfcN. Proteins belonging to this subfamily exist in both prokaryotes and eukaryotes. In general, MutS2 and the stand-alone type Smr domains do not coexist within the same organism, except for a few species [8, 11]. We refer to the stand-alone type Smr domain as family 3 Smr domains (Figure 1(c)). The amino acid sequence alignments revealed some differences between these 3 Smr subfamilies (Figures 1(a), 1(b), and 1(c)). The HGXG motif (underlined in Figures 1(a) and 1(c)) is characteristic for family 1 and 3 Smr domains. On the other hand, family 2 Smr domains contain a TGXG motif (underlined in Figure 1(b)) rather than the HGXG motif



(a) Family 1 Smr domains

(b) Family 2 Smr domains

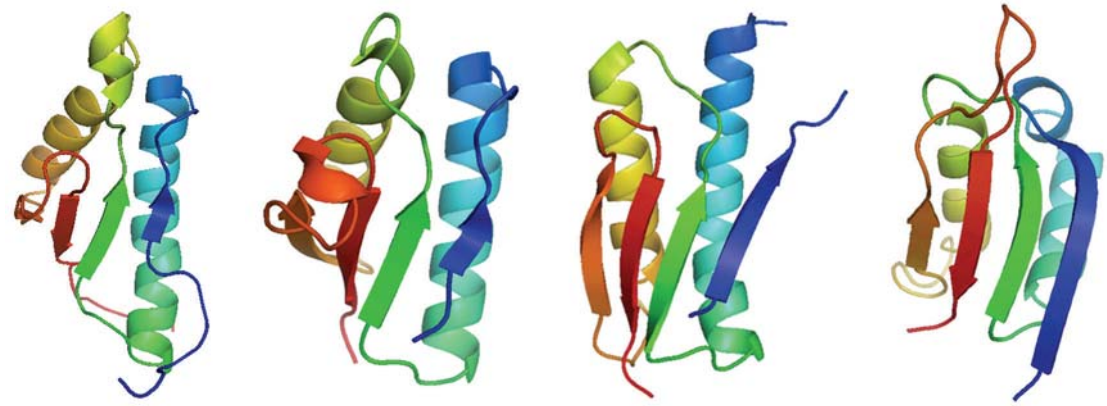
(c) Family 3 Smr domains

Figure 1: Amino acid sequence alignments of family 1, 2, and 3 Smr domains. (a) Family 1 Smr domains. The schematic representation of *T. thermophilus* MutS2 is shown at the top of the panel. The location of the secondary structure elements of *T. thermophilus* MutS2 Smr domain is shown above the sequences. Glu-677 and His-701 in *T. thermophilus* MutS2 are indicated by numbers above the sequences. *Red* and *blue* characters indicate residues whose chemical characteristics are conserved in all and 5 of the 6 species, respectively. The highly conserved HGKG motif is underlined. The 100% conserved acidic residues are indicated with asterisks. (b) Family 2 Smr domains. The schematic representation of *H. sapiens* N4BP2 is shown at the top of the panel. The location of the secondary structure elements of the Smr domain of *H. sapiens* N4BP2 is shown above the sequences. Lys-1722, Ser-1735, Arg-1741, Lys-1743, and Arg-1756 are indicated by numbers above the sequences. *Red* and *blue* characters indicate residues whose chemical characteristics are conserved in all and 4 of the 5 species, respectively. The 100% conserved acidic residues are indicated with asterisks. (c) Family 3 Smr domains. The schematic representation of *E. coli* YdaL is shown at the top of the panel. The location of the secondary structure elements of *E. coli* YdaL is shown above the sequences. *Red* and *blue* backgrounds indicate residues whose chemical characteristics are conserved in all and 4 of the 5 species, respectively. Perfectly conserved basic residues are indicated with asterisks. The 100% conserved acidic residues are indicated by asterisks.

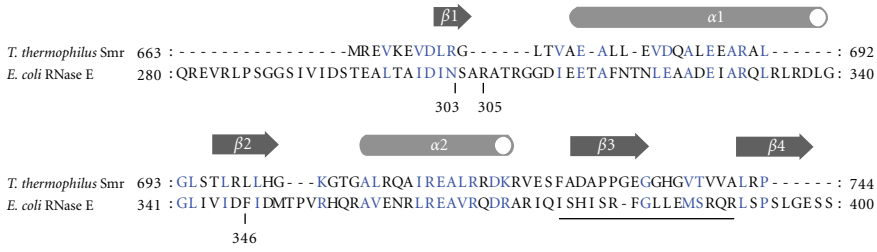
and a highly conserved LDXH motif in their N-terminal regions (underlined in Figure 1(b)). It should be mentioned that motifs somewhat similar to the LDXH are also found in family 1 and 3 Smr domains, implying the significance of this region in the function of Smr domains.

3. Structures of the 3 Smr Subfamilies

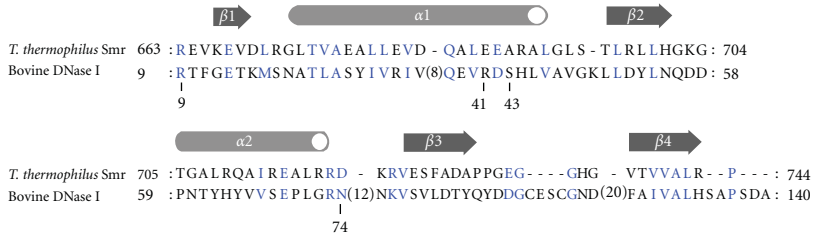
The solution structure of a family 2 Smr domain, namely, the C-terminal domain of human N4BP2 has been solved (PDB ID: 2D9I (unpublished), and 2VKC [19]). The overall structure of the human N4BP2 C-terminal domain comprises an α/β-sandwich structure with a $\beta\alpha\beta\alpha\beta\beta$ fold consisting of a four-stranded β-sheet stacked against 2 α-helices (Figure 2(a)). As shown in Figure 2(b), the crystal structure of a family 1 Smr domain, the *T. thermophilus* MutS2 C-terminal domain (2ZQE), has also been determined, and it shows the same overall structure as the human N4BP C-terminal domain (Z-score: 10.65; root mean square deviation (r.m.s.d.): 2.2 Å; sequence identity: 27%) [14]. Although the coordinate file has not yet been released, the crystal structure of a family 3 Smr domain, residues 39–175 of *E. coli* YdaL, was reported [20]. The overall structure of residues 86–170 of *E. coli* YdaL is also similar to those of the C-terminal domains



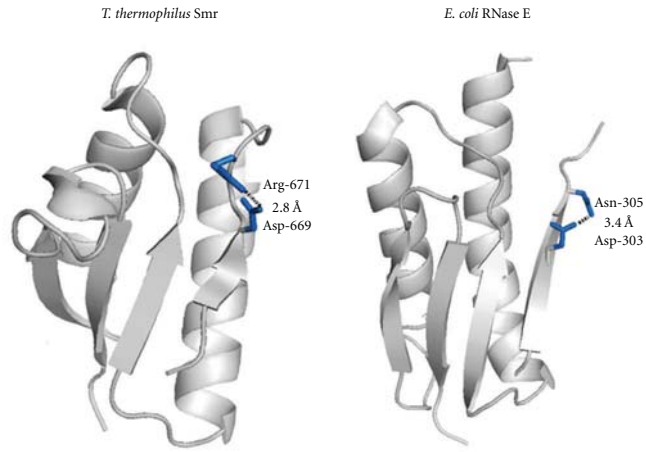
(a) Family 2 Smr domain (b) Family 1 Smr domain (c) *E. coli* RNase E (d) Bovine DNase I



(e)



(f)



(g)

Figure 2: Three-dimensional structures of Smr domains. (a) Solution structure of human N4BP2 Smr domain, a family 2 Smr domain (2d9i). (b) Crystal structure of *T. thermophilus* MutS2 Smr domain, a family 1 Smr domain (2zqe). (c) Crystal structure of the catalytic domain of *E. coli* RNase E (2c0b) [22]. (d) Crystal structure of the N-terminal domain of bovine DNase I (1dnk) [23]. (e) Amino acid sequence comparison between *T. thermophilus* MutS2 Smr domain and the catalytic domain of *E. coli* RNase E. The location of the secondary structure elements of the Smr domain is shown above the sequence. The residues essential for the catalytic activity of RNase E are shown by numbers below the sequence. The dimeric interface in the *E. coli* RNase E catalytic domain is underlined. (f) Amino acid sequence comparison between *T. thermophilus* MutS2 Smr domain and the N-terminal domain of bovine DNase I. The location of the secondary structure elements of the Smr domain is shown above the sequence. The residues essential for the DNA-binding ability are shown by numbers below the sequence. (g) *Left*: a salt bridge between Asp-669 and Arg-671 in the *T. thermophilus* Smr domain is represented with a broken line. *Right*: A hydrogen bond between Asp-303 and Asn-305 is also represented with a broken line.

of human N4BP2 and *T. thermophilus* MutS2. Thus, all 3 Smr subfamilies share a common overall structure as expected from their sequence similarity. As pointed out by Gui et al. [20], the largest structural difference among the 3 Smr subfamilies was found in the length of Loop3 that contains the highly conserved HGXG/TGXG motif. The family 2 Smr domain has the longest loop, while the family 1 Smr domain has the shortest one.

Structural homology search program suggested similarity between Smr domains and a variety of proteins including nucleotide-binding proteins [19]. Interestingly, the overall structure of *T. thermophilus* MutS2 Smr domains shows homology to those of the catalytic domain of *E. coli* RNase E (2BX2) (Z-score: 6.63; r.m.s.d: 2.9 Å; sequence identity: 11%) and the N-terminal DNA-binding domain of bovine DNase I (2DNJ) (Z-score: 3.4; r.m.s.d.: 3.0 Å; sequence identity: 10%) (Figures 2(c) and 2(d)). Diercks et al. [19] also reported that the human N4BP2 C-terminal domain has a tertiary structure homologous to that of the N-terminal domain of bovine DNase I (2DNJ) (Z-score: 4.5). The primary structure comparisons between the *T. thermophilus* MutS2 Smr domain and the catalytic domain of *E. coli* RNase E or the N-terminal domain of bovine DNase I are shown in Figures 2(e) and 2(f), respectively. Although the HGXG/TGXG motif in loop 3 is not found in *E. coli* RNase E and bovine DNase I, some of the residues required for the catalytic activity of RNase E or for the DNA-binding activity of DNase I seem to be conserved in the Smr domain (Figures 2(e) and 2(f)). Interestingly, bovine DNase I has a C-terminal catalytic domain whose primary and tertiary structures are similar to those of its N-terminal domain [23]. It has been pointed out that the C-terminal catalytic domain of bovine DNase I also shows slight structural similarity to the Smr domain [19].

In addition to *E. coli* RNase E and bovine DNase I, a variety of proteins reveal a structural similarity to the Smr domains. For example, a similarity between the human N4BP2 Smr domain and the following proteins was reported [19]: *E. coli* YhhP, a putative cell division protein with an RNA-binding activity (1DCJ) [24] (Z-score: 7.4), the C-terminal domain of *Bacillus stearothermophilus* IF3C, a translational initiation factor (1TIG) [25] (Z-score: 6.0), the N-terminal subdomain of the *B. stearothermophilus* ribosomal S8 protein (1SEI) [26] (Z-score: 5.9), and the R3H domain, a putative single-stranded nucleic acid-binding domain of human Smubp-2 (1MSZ) [27] (Z-score: 4.9). Thus, Smr domains share a widely conserved fold with various kinds of oligonucleotide-binding proteins.

4. All 3 Subfamilies of Smr Domains Have Endonuclease Activity

As mentioned above, the hypothesis that the Smr domain has an endonuclease activity has been proposed on the basis of the domain-architecture analogy of MutS2 to another endonuclease domain-containing MutS homologue [17]. The first experimental evidence for the endonuclease activity of the Smr domain was provided by the functional characterization of the C-terminal domain of human N4BP2, a family 2 Smr domain. The recombinant C-terminal domain of human N4BP2 incised a supercoiled plasmid DNA to generate an open circular form of the plasmid, demonstrating the nicking endonuclease activity of this protein [19, 21]. Next, the endonuclease activity of family 1 Smr domain was also confirmed. The Smr domain of *T. thermophilus* MutS2 was shown to relax supercoiled plasmid DNA and digest linear double-stranded DNA [28, 29]. Finally, it was demonstrated that *E. coli* YdaL, a family 3 Smr domain, exhibits endonuclease activity against supercoiled plasmid DNA [20]. Thus, all of 3 subfamilies of Smr domain have been verified to be endonuclease domains.

The k_{cat} and K_M values for the endonuclease activity of the family 1 Smr domain against linear double-stranded DNA have been reported to be 0.041 min^{-1} and 290 nM, respectively [18]. The k_{cat} value is at least 20 times higher than that for a DNA mismatch repair nicking endonuclease MutL [30]. However, it would be appropriate to avoid kinetic parameter-based discussions until the parameters are determined using the most favorable substrate for each Smr domain.

In order to elucidate the mechanism by which the Smr domain incises DNA, the products generated by Smr reaction were analyzed by mass spectrometry and the chemical nature of the cleaved DNA termini was identified. The result clearly showed that products of the Smr reaction contain 3′-hydroxy and 5′-phosphate termini, indicating that Smr hydrolyzes the phosphodiester bond of the deoxyoligonucleotides at 5′-side of the phosphate [31]. It has been also shown that the Smr domains require divalent metal cations for the reaction [19, 20, 29]. Generally, divalent metal ion-dependent nucleases require acidic amino acid residues to coordinate the metal ions [32, 33]. However, the catalytic residues of the Smr domain have not yet been identified to date.

The pH dependence of the k_{cat} value of the *T. thermophilus* MutS2 Smr domain suggested that the endonuclease activity depends on the basic form of the amino acid side chain, which has a pK_a value around 6 [18]. In *T. thermo-philus* MutS2, substitution of His-701 with alanine resulted in a drastic decrease in the velocity of the activity [18]. His-701 of the *T. thermophilus* MutS2 is within the HGXG sequence motif that is conserved in the family 1 and 3 Smr domains (Figure 1(a)). Nevertheless, the histidine residue in the HGXG motif is not conserved in the family 2 Smr domain. In addition, residual activity of H701A mutant of *T. thermophilus* MutS2 was observed [29], implying the involvement of other amino acid residues in the catalysis.

The structural homology of Smr domains to *E. coli* RNase E and bovine DNase I may provide the clue to explore the catalytic residues of Smr domains. Asp-303 and Asp-346 in *E. coli* RNase E are the catalytic residue coordinating a magnesium ion, and Asn-305 supports the orientation of Asp-303 via hydrogen bonding [22] (Figure 2(e)). The majority of Smr domains contain an aspartate at the site (Asp-669 and Asp-1692 in the *T. thermophilus* MutS2 and *H. sapiens* N4BP2 Smr domains, resp.) that spatially corresponds to Asp-303 of *E. coli* RNase E (Figures 1 and 2(e)). Furthermore, the orientation of Asp-669 in the *T. thermophilus* MutS2 Smr

domain is adjusted by a salt bridge with Arg-671 [14], which is located in the site corresponding to Asn-305 of *E. coli* RNase E (Figure 2(g)). The involvement of those aspartate residues may be suspicious; however, *E. coli* YdaL lacks this aspartate residue.

Primary structure comparison revealed that Glu-677 in *T. thermophilus* MutS2 shows relatively high level of conservation among all 3 Smr subfamilies [11], implying the possible involvement of this acidic residue as a catalytic residue. However, *E. coli* YdaL also does not have an acidic residue at the corresponding site. It would be possible that 3 subfamilies employ different amino acid residues to catalyze the reaction.

The residues required for the DNA-binding ability of the family 2 Smr domain (human N4BP2 Smr domain) were surveyed in detail by using NMR measurement and site-directed mutagenesis [19]. The residues whose chemical shifts were affected by addition of bubble DNA structure are mapped to loops 1, 3, 4, and 5 of the human N4BP2 Smr domain. Subsequent site-directed mutagenesis confirmed the significant requirements of the basic and neutral residues in loops 2 (Lys-1722), 3 (Ser-1735, Arg-1741, and Lys-1743), and 4 (Arg-1756) for DNA-binding activity. Since the residues around loop 4 are poorly conserved, Diercks et al. [19] discussed the possibility that DNA-binding induced secondary effects on the local structures of these residues. While Arg-1741 and Lys-1743 are relatively conserved in all 3 Smr subfamilies, Lys-1722 seems to exist only in family 2, and Ser-1735 is conserved only in families 1 and 2 (Figure 1). It should be noted that Arg-1741 and Ser-1735 are located near the site spatially corresponding to the DNA-binding residues of bovine DNase I [19]. The DNA-binding mode of a single polypeptide of the Smr domain may be analogous to that of DNase I. However, as discussed later, we should take into account the quaternary structure of the Smr domains when we consider their DNA-binding mode.

5. Substrate Specificity of Smr Domains

It has been reported that MutS2 preferably binds to branched DNA structures, such as Holliday junctions, D-loops, and pseudo-Y structures [10, 11, 14]. Its binding specificity for branched DNA structures is analogous to that of another MutS paralogue, MutSγ (that comprises MSH4 and MSH5) [34, 35], which does not contain Smr domain. Therefore, the involvement of the Smr domain in the recognition of branched DNA structures was unexpected. In fact, the binding affinity of the Smr-deleted mutant (N-terminal domain) of *T. thermophilus* MutS2 is as tight as that of the intact MutS2 [14]. However, surprisingly, the Smr domain of *T. thermophilus* MutS2 showed specificity to the Holliday junction though its K_d value (260 nM) was significantly higher than that of the N-terminal domain (60 nM) [14]. In addition, it was also reported that the human N4BP2 Smr domain and the *E. coli* YdaL domain showed a significant binding preference for branched DNA structures, including bubble DNA structure and Holliday junctions [14, 19]. Substrate specificity for branched DNA structures is a common feature among all 3 Smr subfamilies.

Generally, branched DNA-recognizing proteins, with a few exceptions, are dimeric or tetrameric molecules, because they need to hold multiple "arms" of the substrate. For instance, T7 endonuclease I [36], T4 endonuclease VII [37], eukaryotic MutSγ (MSH4/MSH5) [34], eukaryotic MUS81-EME1 [38], and archaeal Hef [39] function in dimeric forms, and bacterial RuvA [40] and RuvC [41] are known to be tetrameric. Therefore, it is possible that Smr domains are in an oligomeric state in their functional form, although the three-dimensional structure of oligomerized Smr domains has not yet been reported. As to family 1 Smr domain, 621–662 residues of *T. thermophilus* MutS2, which are located between the N-terminal and the Smr domains, are responsible for the dimerization of the Smr domain [14, 18]. Consistent with this, a family 2 Smr domain, the human N4BP2 Smr domain, also forms a dimeric molecule upon DNA binding [19]. Recently, the DNA-binding and endonuclease activities of a family 3 Smr, *E. coli* YdaL, were found to be enhanced by the presence of the N-terminal 1–38 residues that are not included in Smr core domain [20]. Those N-terminal residues may affect the oligomeric state of *E. coli* YdaL. It has been known that *E. coli* RNase E also functions in a dimeric form [22] and that its dimer interface is located in the core region of the catalytic domain (the underlined region in Figure 2(b)), suggesting that the quaternary structures of the Smr domain and the *E. coli* RNase E catalytic domain are quite dissimilar to each other. The quaternary structure is likely to be closely correlated to the substrate specificity of these widely distributed folds of proteins.

6. Molecular and Cellular Functions of Smr-Containing Proteins

As mentioned above, a variety of Smr-containing proteins are distributed across a wide range of organisms. The biochemical characterization of Smr domains would improve the understanding of their cellular functions.

The family 1 Smr-containing protein MutS2 has been implicated to participate not only in the suppression of homologous recombination but also in the protection of cells from oxidative DNA damages [42, 43]. *Helicobacter pylori* MutS2 recognizes DNA containing 8-oxoguanine, a major DNA lesion caused by oxidative stress, and deletion of *mutS2* gene results in an accumulation of 8-oxoguanine in the cell [43]. Endonuclease activities are often required for DNA repair pathway to conduct the downstream excision reactions of damaged nucleotides [2]. It would be intriguing to test the activity of the Smr domain on 8-oxoguanine-containing DNA.

Family 2 Smr-containing proteins are extreme diverse in their domain architecture. Among these, the plant GUN1 and the mammalian N4BP2 are relatively well characterized. *Arabidopsis thaliana* GUN1 was identified as the key component in the plastid-to-nucleus retrograde signaling

pathways that couple nuclear gene expressions and chloroplast functions [44]. GUN1, a member of pentatricopeptide repeat- (PPR-) containing proteins, has a Smr domain in its C-terminal region. Most of the PPR-containing proteins are thought to function in processing and stabilizing RNA molecules [45, 46], as well as in interacting with DNA molecules [47]. It was confirmed that the GUN1 Smr domain binds to DNA, and its binding activity was affected by the PPR motif [44]. Further experiments revealed that on plastid DNA, GUN1 is located at the sites that are being actively transcribed [44]. In addition to GUN1, *Arabidopsis thaliana* has at least 7 GUN1 paralogues that contain both the PPR motif and the Smr domain [48]. One of them, pTAC2, has been reported to colocalize with GUN1 at the site of actively transcribed plastid and thought to be responsible for the plastid gene expression [48, 49]. Another GUN1 paralogue, SVR7, has been discussed to be directly involved in chloroplast rRNA processing [48], where the endonuclease activity may be required.

Mammalian N4BP2 was originally identified as a protein that specifically interacts with the E3 ubiquitin ligase NEDD4 [50]. Subsequent studies also revealed specific interaction of N4BP2 with BCL3 [21]. BCL3 is thought to activate transcription by interacting with transcription factors and other DNA-binding proteins [51, 52]. Induction of both NEDD4 and BCL3 is known to be correlated with various types of cancer including human breast cancer [53, 54], and the association of N4BP2 itself with sporadic carcinoma has also been reported [55]. It remains to be investigated whether the branched DNA-specific binding and/or the endonuclease activity of the Smr domain is involved in the transcription-regulatory role of N4BP2. It should be mentioned that a highly conserved domain of unknown function DUF1771 (in Pfam [56]) is often adjacent to family 2 Smr-domain in eukaryotes (Table 1). The Smr domains in mammalian N4BP2 are also accompanied by DUF1771. Structural and functional analyses of DUF1771 would provide information important for illustrating the molecular function of family 2 Smr domains.

Lactobacillus casei phage ϕFSW repressor can be classified as a family 3 Smr domain [16]. It would be worth elucidating whether the repressor protein has endonuclease activity. The ϕFSW repressor protein has no N-terminal stretch, and its molecular and cellular function may be distinct from those of other family 3 Smr domains. To date, there are no reports about the cellular functions of other stand-alone type Smr domains.

The transcription regulatory role of the ϕFSW repressor protein is reminiscent of those of GUN1 and N4BP2. Although an endonuclease activity hardly seems to be correlated with the regulation of transcription, it has been shown that human NM23-H2 is a transcriptional regulator with DNA-cleaving activity [57, 58]. Furthermore, it has been clarified that mammalian nucleotide excision repair (NER) components including XPG and ERCC1-XPF endonucleases are recruited to the transcription machinery at the promoter of nuclear receptor genes [59]. NER is known to function in transcription-coupled repair, which rescues the stalled RNA polymerase II by repairing DNA lesions and requires CSB protein as a mediator [60, 61]. However, the recruitment of NER endonucleases to the promoters is independent of the exogenous genotoxic agents and transcription coupling repair-specific CSB [62]. Thus, in addition to transcription-coupled repair of DNA lesion, NER endonucleases may also participate in transcription itself. Le May et al. discussed the possible role of NER components in chromatin remodeling during the transcription [59, 62]. These observations lead us to the supposition that DNA-cleaving activity of Smr domain may play a role in the regulation of transcription. There is, of course, another possibility that ϕFSW repressor protein, GUN1, and N4BP2 have multiple cellular functions and the endonuclease activity is not correlated with the regulation of transcription.

7. Conclusions

The sequences homologous to the Smr domains of MutS2 proteins are conserved in almost all organisms except for archaea. Smr domains are classified into 3 subfamilies on the basis of the domain architecture of the proteins in which Smr domains are present. Three-dimensional structures of Smr domains revealed that all 3 subfamilies share a common overall structure despite the local differences in loop regions. Consistent with this, all 3 subfamilies showed endonuclease activity and specificity for branched DNA structures. Immediate identification of the catalytic residues is required to study the reaction mechanism of this endonuclease. Since the relationship between the cellular and molecular functions of the majority of the family 2 or 3 Smr domain-containing proteins is still unknown, detailed characterization of these Smr domains may lead to the discovery of a novel biological phenomenon. For this purpose, an unavoidable task in the future will be to identify the most preferable substrate of the endonuclease or DNA-binding activity.

Acknowledgments

The authors thank Drs. Ryoji Masui and Noriko Nakagawa for their valuable discussions on this study.

References

[1] E. C. Friedberg, G. C. Walker, R. D. Wood, R. A. Schultz, and T. Ellenberger, *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, DC, USA, 2nd edition, 2006.

[2] R. Morita, S. Nakane, A. Shimada et al., "Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems," *Journal of Nucleic Acids*, vol. 2010, Article ID 179594, 2010.

[3] T. A. Kunkel and D. A. Erie, "DNA mismatch repair," *Annual Review of Biochemistry*, vol. 74, pp. 681–710, 2005.

[4] P. Modrich, "Mechanisms in eukaryotic mismatch repair," *Journal of Biological Chemistry*, vol. 281, no. 41, pp. 30305–30309, 2006.

[5] K. Fukui, "DNA mismatch repair in eukaryotes and bacteria," *Journal of Nucleic Acids*, vol. 2010, Article ID 260512, 16 pages, 2010.

[6] J. A. Eisen and P. C. Hanawalt, "A phylogenomic study of DNA repair genes, proteins, and processes," *Mutation Research*, vol. 435, no. 3, pp. 171–213, 1999.

[7] Z. Lin, M. Nei, and H. Ma, "The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution," *Nucleic Acids Research*, vol. 35, no. 22, pp. 7591–7603, 2007.

[8] P. Sachadyn, "Conservation and diversity of MutS proteins," *Mutation Research*, vol. 694, no. 1-2, pp. 20–30, 2010.

[9] H. Ogata, J. Ray, K. Toyoda et al., "Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment," *ISME Journal*, vol. 5, no. 7, pp. 1143–1151, 2011.

[10] A. V. Pinto, A. Mathieu, S. Marsin et al., "Suppression of homologous and homeologous recombination by the bacterial MutS2 protein," *Molecular Cell*, vol. 17, no. 1, pp. 113–120, 2005.

[11] J. Kang, S. Huang, and M. J. Blaser, "Structural and functional divergence of MutS2 from bacterial MutS1 and eukaryotic MSH4-MSH5 homologst," *Journal of Bacteriology*, vol. 187, no. 10, pp. 3528–3537, 2005.

[12] P. Rossolillo and A. M. Albertini, "Functional analysis of the *Bacillus subtilis yshD* gene, a mutS paralogue," *Molecular and General Genetics*, vol. 264, no. 6, pp. 809–818, 2001.

[13] S. Mennecier, G. Coste, P. Servant, A. Bailone, and S. Sommer, "Mismatch repair ensures fidelity of replication and recombination in the radioresistant organism *Deinococcus radiodurans*," *Molecular Genetics and Genomics*, vol. 272, no. 4, pp. 460–469, 2004.

[14] K. Fukui, N. Nakagawa, Y. Kitamura, Y. Nishida, R. Masui, and S. Kuramitsu, "Crystal structure of Muts2 endonuclease domain and the mechanism of homologous recombination suppression," *Journal of Biological Chemistry*, vol. 283, no. 48, pp. 33417–33427, 2008.

[15] D. Moreira and H. Philippe, "Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family," *Trends in Biochemical Sciences*, vol. 24, no. 8, pp. 298–300, 1999.

[16] B. Binishofer, I. Moll, B. Henrich, and U. Bläsi, "Inducible promoter-repressor system from the Lactobacillus casei phage *φ*FSW," *Applied and Environmental Microbiology*, vol. 68, no. 8, pp. 4132–4135, 2002.

[17] H. S. Malik and S. Henikoff, "Dual recognition-incision enzymes might be involved in mismatch repair and meiosis," *Trends in Biochemical Sciences*, vol. 25, no. 9, pp. 414–418, 2000.

[18] K. Fukui, H. Kosaka, S. Kuramitsu, and R. Masui, "Nuclease activity of the MutS homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain," *Nucleic Acids Research*, vol. 35, no. 3, pp. 850–860, 2007.

[19] T. Diercks, E. Ab, M. A. Daniels et al., "Solution structure and characterization of the DNA-binding activity of the B3BP-Smr domain," *Journal of Molecular Biology*, vol. 383, no. 5, pp. 1156–1170, 2008.

[20] W.-J. Gui, Q.-H. Qu, Y.-Y. Chen et al., "Crystal structure of YdaL, a stand-alone small MutS-related protein from *Escherichia coli*," *Journal of Structural Biology*, vol. 174, no. 2, pp. 282–289, 2011.

[21] N. Watanabe, S. Wachi, and T. Fujita, "Identification and characterization of BCL-3-binding protein: implications for transcription and DNA repair or recombination," *Journal of Biological Chemistry*, vol. 278, no. 28, pp. 26102–26110, 2003.

[22] A. J. Callaghan, M. J. Marcaida, J. A. Stead, K. J. McDowall, W. G. Scott, and B. F. Luisi, "Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover," *Nature*, vol. 437, no. 7062, pp. 1187–1191, 2005.

[23] S. A. Weston, A. Lahm, and D. Suck, "X-ray structure of the DNase I-d(GGTATACC) complex at 2.3 Å resolution," *Journal of Molecular Biology*, vol. 226, no. 4, pp. 1237–1256, 1992.

[24] E. Katoh, T. Hatta, H. Shindo et al., "High precision NMR structure of YhhP, a novel *Escherichia coli* protein implicated in cell division," *Journal of Molecular Biology*, vol. 304, no. 2, pp. 219–229, 2000.

[25] V. Biou, F. Shu, and V. Ramakrishnan, "X-ray crystallography shows that translational initiation factor IF3 consists of two compact α/β domains linked by an α-helix," *EMBO Journal*, vol. 14, no. 16, pp. 4056–4064, 1995.

[26] C. Davies, V. Ramakrishnan, and S. W. White, "Structural evidence for specific S8-RNA and S8-protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 Å resolution," *Structure*, vol. 4, no. 9, pp. 1093–1104, 1996.

[27] E. Liepinsh, A. Leonchiks, A. Sharipo, L. Guignard, and G. Otting, "Solution structure of the R3H domain from human Sμbp-2," *Journal of Molecular Biology*, vol. 326, no. 1, pp. 217–223, 2003.

[28] K. Fukui, R. Masui, and S. Kuramitsu, "*Thermus thermophilus* MutS2, a MutS paralogue, possesses an endonuclease activity promoted by MutL," *Journal of Biochemistry*, vol. 135, no. 3, pp. 375–384, 2004.

[29] K. Fukui, H. Kosaka, S. Kuramitsu, and R. Masui, "Nuclease activity of the MutS homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain," *Nucleic Acids Research*, vol. 35, no. 3, pp. 850–860, 2007.

[30] J. Mauris and T. C. Evans, "Adenosine triphosphate stimulates *Aquifex aeolicus* MutL endonuclease activity," *PLoS ONE*, vol. 4, no. 9, Article ID e7175, 2009.

[31] K. Fukui, Y. Takahata, N. Nakagawa, S. Kuramitsu, and R. Masui, "Analysis of a nuclease activity of catalytic domain of *Thermus thermophilus* MutS2 by high-accuracy mass spectrometry," *Nucleic Acids Research*, vol. 35, no. 15, Article ID e100, 2007.

[32] W. Yang, "An equivalent metal ion in one- and two-metal-ion catalysis," *Nature Structural and Molecular Biology*, vol. 15, no. 11, pp. 1228–1231, 2008.

[33] W. Yang, "Nucleases: diversity of structure, function and mechanism," *Quarterly Reviews of Biophysics*, vol. 44, no. 1, pp. 1–93, 2011.

[34] T. Snowden, S. Acharya, C. Butz, M. Berardini, and R. Fishel, "hMSH4-hMSH5 recognizes holliday junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes," *Molecular Cell*, vol. 15, no. 3, pp. 437–451, 2004.

[35] N. M. Hollingsworth, L. Ponte, and C. Halsey, "MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair," *Genes and Development*, vol. 9, no. 14, pp. 1728–1739, 1995.

[36] J. M. Hadden, A. C. Déclais, S. B. Carr, D. M. J. Lilley, and S. E. V. Phillips, "The structural basis of Holliday junction resolution by T7 endonuclease I," *Nature*, vol. 449, no. 7162, pp. 621–624, 2007.

[37] C. Biertümpfel, W. Yang, and D. Suck, "Crystal structure of T4 endonuclease VII resolving a Holliday junction," *Nature*, vol. 449, no. 7162, pp. 616–620, 2007.

[38] H. C. Jeong, J. K. Jeong, M. C. Jung, H. L. Jung, and Y. Cho, "Crystal structure of the Mus81-Eme1 complex," *Genes and Development*, vol. 22, no. 8, pp. 1093–1106, 2008.

[39] T. Nishino, K. Komori, D. Tsuchiya, Y. Ishino, and K. Morikawa, "Crystal structure and functional implications of Pyrococcus furiosus Hef helicase domain involved in branched DNA processing," *Structure*, vol. 13, no. 1, pp. 143–153, 2005.

[40] M. Ariyoshi, T. Nishino, H. Iwasaki, H. Shinagawa, and K. Morikawa, "Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 15, pp. 8257–8262, 2000.

[41] M. Ariyoshi, D. G. Vassylyev, H. Iwasaki, H. Nakamura, H. Shinagawa, and K. Morikawa, "Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*," *Cell*, vol. 78, no. 6, pp. 1063–1072, 1994.

[42] G. Wang, P. Alamuri, and R. J. Maier, "The diverse antioxidant systems of *Helicobacter pylori*," *Molecular Microbiology*, vol. 61, no. 4, pp. 847–860, 2006.

[43] G. Wang, P. Alamuri, M. Zafri Humayun, D. E. Taylor, and R. J. Maier, "The *Helicobacter pylori* MutS protein confers protection from oxidative DNA damage," *Molecular Microbiology*, vol. 58, no. 1, pp. 166–176, 2005.

[44] S. Koussevitzky, A. Nott, T. C. Mockler et al., "Signals from chloroplasts converge to regulate nuclear gene expression," *Science*, vol. 316, no. 5825, pp. 715–719, 2007.

[45] C. Lurin, C. Andrés, S. Aubourg et al., "Genome-wide analysis of arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis," *Plant Cell*, vol. 16, no. 8, pp. 2089–2103, 2004.

[46] C. Schmitz-Linneweber, R. E. Williams-Carrier, P. M. Williams-Voelker, T. S. Kroeger, A. Vichas, and A. Barkan, "A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA," *Plant Cell*, vol. 18, no. 10, pp. 2650–2663, 2006.

[47] N. Tsuchiya, H. Fukuda, K. Nakashima, M. Nagao, T. Sugimura, and H. Nakagama, "LRP130, a single-stranded DNA/RNA-binding protein, localizes at the outer nuclear and endoplasmic reticulum membrane, and interacts with mRNA in vivo," *Biochemical and Biophysical Research Communications*, vol. 317, no. 3, pp. 736–743, 2004.

[48] X. Liu, F. Yu, and S. Rodermel, "An arabidopsis pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis," *Plant Physiology*, vol. 154, no. 4, pp. 1588–1601, 2010.

[49] J. Pfalz, K. Liere, A. Kandlbinder, K. J. Dietz, and R. Oelmüller, "pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression," *Plant Cell*, vol. 18, no. 1, pp. 176–197, 2006.

[50] K. F. Harvey and S. Kumar, "Nedd4-like proteins: an emerging family of ubiquitin-protein ligases implicated in diverse cellular functions," *Trends in Cell Biology*, vol. 9, no. 5, p. 166, 1999.

[51] S. Y. Na, H. S. Choi, J. W. Kim, D. S. Na, and J. W. Lee, "Bc13, an IκB protein, as a novel transcription coactivator of the retinoid X receptor," *Journal of Biological Chemistry*, vol. 273, no. 47, pp. 30933–30938, 1998.

[52] S. Y. Na, J. E. Choi, H. J. Kim, B. H. Jhun, Y. C. Lee, and J. W. Lee, "Bcl3, an IκB protein, stimulates activating protein-1 transactivation and cellular proliferation," *Journal of Biological Chemistry*, vol. 274, no. 40, pp. 28491–28496, 1999.

[53] P. C. Cogswell, D. C. Guttridge, W. K. Funkhouser, and A. S. Baldwin, "Selective activation of NF-κB subunits in human breast cancer: potential roles for NF-κB2/p52 and for Bcl-3," *Oncogene*, vol. 19, no. 9, pp. 1123–1131, 2000.

[54] C. Chen and L. E. Matesic, "The Nedd4-like family of E3 ubiquitin ligases and cancer," *Cancer and Metastasis Reviews*, vol. 26, no. 3-4, pp. 587–604, 2007.

[55] M. Z. Zheng, H. D. Qin, X. J. Yu et al., "Haplotype of gene Nedd4 binding protein 2 associated with sporadic nasopharyngeal carcinoma in the Southern Chinese population," *Journal of Translational Medicine*, vol. 5, Article ID 36, 2007.

[56] R. D. Finn, J. Mistry, J. Tate et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 38, supplement 1, pp. D211–D222, 2009.

[57] E. H. Postel, B. A. Abramczyk, S. K. Gursky, and Y. Xu, "Structure-based mutational and functional analysis identify human NM23-H2 as a multifunctional enzyme," *Biochemistry*, vol. 41, no. 20, pp. 6330–6337, 2002.

[58] E. H. Postel, B. M. Abramczyk, M. N. Levit, and S. Kyin, "Catalysis of DNA cleavage and nucleoside triphosphate synthesis by NM23-H2/NDP kinase share an active site that implies a DNA repair function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 26, pp. 14194–14199, 2000.

[59] N. Le May, J.-M. Egly, and F. Coin, "True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair," *Journal of Nucleic Acids*, vol. 2010, Article ID 616342, 10 pages, 2010.

[60] M. Fousteri and L. H. F. Mullenders, "Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects," *Cell Research*, vol. 18, no. 1, pp. 73–84, 2008.

[61] P. C. Hanawalt and G. Spivak, "Transcription-coupled DNA repair: two decades of progress and surprises," *Nature Reviews Molecular Cell Biology*, vol. 9, no. 12, pp. 958–970, 2008.

[62] N. Le May, D. Mota-Fernandes, R. Vélez-Cruz, I. Iltis, D. Biard, and J. M. Egly, "NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack," *Molecular Cell*, vol. 38, no. 1, pp. 54–66, 2010.

SAGE-Hindawi Access to Research
Molecular Biology International
Volume 2011, Article ID 475718, 9 pages
doi:10.4061/2011/475718

Review Article

Databases and Bioinformatics Tools for the Study of DNA Repair

Kaja Milanowska,[1,2] Kristian Rother,[1,2] and Janusz M. Bujnicki[1,2]

[1] *Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland*
[2] *Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznan, Poland*

Correspondence should be addressed to Janusz M. Bujnicki, iamb@genesilico.pl

Received 16 February 2011; Revised 28 April 2011; Accepted 22 May 2011

Academic Editor: Frédéric Coin

DNA is continuously exposed to many different damaging agents such as environmental chemicals, UV light, ionizing radiation, and reactive cellular metabolites. DNA lesions can result in different phenotypical consequences ranging from a number of diseases, including cancer, to cellular malfunction, cell death, or aging. To counteract the deleterious effects of DNA damage, cells have developed various repair systems, including biochemical pathways responsible for the removal of single-strand lesions such as base excision repair (BER) and nucleotide excision repair (NER) or specialized polymerases temporarily taking over lesion-arrested DNA polymerases during the S phase in translesion synthesis (TLS). There are also other mechanisms of DNA repair such as homologous recombination repair (HRR), nonhomologous end-joining repair (NHEJ), or DNA damage response system (DDR). This paper reviews bioinformatics resources specialized in disseminating information about DNA repair pathways, proteins involved in repair mechanisms, damaging agents, and DNA lesions.

1. Introduction

DNA repair processes are indispensable for maintaining the integrity of genetic information in all organisms. Environmental agents such as chemicals, UV light, and ionizing radiation, as well as endogenous metabolic processes involving DNA constantly challenge the chemical structure and stability of the genome. DNA lesions can interfere with processes such as DNA replication or transcription and may lead to mutations and cancer [1, 2]. To prevent the erosion of the chemical structure of DNA, living systems have evolved various different biochemical systems for DNA repair [3–7].

DNA damage from endogenous sources gives rise to 20,000 lesions per mammalian cell per day. Amongst these lesions, the most common are base deamination, spontaneous hydrolysis of the *N*-glycosidic bond, alkylation, and damage by reactive oxygen or nitrogen species and lipid peroxidation products [8–12]. Other lesions such as the formation of single- and double-strand breaks, the collapse of replication forks, and the introduction of modified nucleic acid bases during DNA replication are caused by errors in DNA metabolic processes. In total, there are 10^{16}–10^{18} DNA repair events that occur daily in a healthy adult man (10^{12} cells) [13]. Lesions that are not repaired often lead to mutations, aging and various diseases, including carcinogenesis and neurodegeneration [14–18]. Some pathological disorders directly related to defects in the DNA repair machinery are Xeroderma pigmentosum, different types of cancer (breast cancer, colorectal cancer, endometrial cancer, gastric cancer, or prostate cancer), Fancomi anemia, Muir-Torre syndrome, Tay syndrome, and Werner syndrome. On the other hand, unrepaired lesions that occur in germline cells become the main source of genetic variability and therefore a driving force for the evolution. For this reason, the DNA repair system needs not only to be regulated to maintain an individual genome's integrity, but also to increase the genetic variability in the context of populations. Many mechanisms are known that regulate the amount of DNA repair as a response to environmental conditions [19].

Given its many duties in different contexts, it is not surprising that DNA repair is a very complicated process, involving many factors. For instance to date, 168 genes

encoding proteins involved in DNA repair have been identified in the human genome [17, 18, 20] (20 January 2011, date last accessed). Over all organisms, there are many more; for base excision repair alone, KEGG [21] lists 41 groups of orthologous genes encoding for hundreds of proteins in total. The key players in DNA repair are enzymes that catalyze reactions leading from the DNA with damage to a repaired molecule. They are assisted by proteins that detect damage and mediate signals that coordinate the repair process with other cellular processes. From the point of view of the DNA substrate, the biochemical pathways of DNA repair can be divided into eight categories:

(i) DNA damage signaling (DDS): also known as the DNA damage checkpoint; it is a group of responses to DNA damage caused by some endogenous and environmental agents; activation of these pathways may be triggered by the effect the DNA lesions have on replication, transcription, or chromatin topology;

(ii) base-excision repair (BER): initiated by excision of a modified base from the DNA. Depending on the length of DNA resynthesis, the pathway is subdivided into two subpathways: short path (SP-BER) or long path (LP-BER);

(iii) DNA damage response (DDR): directly restores the native nucleotide residue by removing the nonnative chemical modification;

(iv) homologous recombination repair (HRR): repair of DNA double-strand breaks using the homologous DNA strand as a template for resynthesis;

(v) mismatch repair (MMR): postreplicational DNA repair that removes errors introduced during the replication (misinserted nucleotides, small loops, insertions, deletions);

(vi) nonhomologous end-joining repair (NHEJ): ligation of ends resulting from DNA double-strand breaks (including the more error-prone microhomology-mediated end-joining (MMEJ) mechanism;

(vii) nucleotide excision repair (NER): removes bulky damage from the DNA. The damage from the active strand of transcribed genes is removed by transcription coupled repair (TCR)-NER, while global genome repair (GGR)-NER removes damage present elsewhere in the genome;

(viii) translesion synthesis (TLS): damage-tolerance pathway that employs specialized polymerases to replicate across lesions in order to finish replication despite DNA damage.

Each of these pathways can be represented as a series of enzymatic transformations between different DNA structures, catalyzed by a dedicated system of proteins. It must be emphasized that DNA repair pathways are connected to each other, that is, they can share some steps and/or proteins involved [13]. As a consequence, DNA repair proteins rarely work in isolation in the cell, and their activity is dependent on other components of DNA repair systems.

DNA repair itself is not an isolated process, and it is strongly connected to other pathways of nucleic acid metabolism, including (but not limited to) DNA replication, DNA epigenetic modification, transcription, cell cycle regulation, and induced cell death as well as processes that are specific to different domains of life, such as telomere maintenance in eukaryotes and DNA restriction in prokaryotes.

2. DNA Repair Data and Databases

The knowledge of DNA repair systems and their components is critical to our understanding of how cells control the integrity of their genomes. A large body of data on this topic has been published mostly in the literature and in a few electronic resources. Today, systematizing this knowledge and presenting it in a clear and easily accessible way is mostly done by biological databases. The collection, curation, and availability of data are necessary to answer questions about subsystems of DNA repair, for example, "which proteins participate in MMR in humans and in plants?", "what immediate cellular response is triggered by damage caused by UV light?", or "how does HRR differ between plants and vertebrates?". The topic of DNA repair is covered by many computational resources. However, there are few databases dedicated to DNA repair, and most of the data is scattered over various general databases. In Table 1, we have listed some of the available web resources relevant to DNA repair, and in the following section we discuss their content.

2.1. Databases Dedicated to DNA Repair. "REPAIRtoire" is a database for systems biology of DNA damage and repair developed by the authors of this paper and their coworkers [22]. The purpose of this database is to gather information about all DNA repair systems and proteins from model organisms and to facilitate the access to knowledge about correlation of human diseases with mutations in genes responsible for DNA integrity and stability as well as information about toxic and mutagenic agents causing DNA damage. REPAIRtoire is available online at http://repairtoire.genesilico.pl/. It organizes data into the following categories: (i) the chemical structures of DNA lesions (as of April 2011: 85 different types of damage in the DNA) linked to their causative mutagenic and cytotoxic agents, (ii) pathways comprising individual processes and enzymatic reactions involved in the removal of damage, (iii) proteins participating in DNA repair, in particular enzymes involved in the transformation between different chemical structures of the DNA substrate, and (iv) diseases correlated with mutations in genes encoding DNA repair proteins (40 diseases caused by the mutations in 32 genes linked to defects in DNA repair proteins). It also provides links to publications and external datasets. REPAIRtoire covers all eight main DNA damage checkpoint, repair, and tolerance pathways (see above). The pathway/protein dataset is currently limited to three model organisms: *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens*. DNA repair and tolerance pathways are represented as graphs and in tabular form with descriptions of each repair step as well as corresponding proteins. The individual entries in the

Table 1: Databases dedicated to DNA repair and general-purpose databases relevant to DNA repair.

Name	url	Reference	Description
Databases dedicated to DNA repair			
REPAIRtoire	http://repairtoire.genesilico.pl/	[22]	Database of DNA repair pathways
repairGENES	http://www.repairgenes.org/	unpublished	Database of DNA repair genes
Human DNA Repair Genes	http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html	[17]	Database of human DNA repair genes
Repair-FunMap	currently unavailable	[23]	A database of interactions between proteins involved in DNA repair and other proteins
Other databases relevant to DNA repair			
KEGG	http://www.genome.jp/kegg/	[21]	Kyoto Encyclopedia of Genes and Genomes
Reactome	http://www.reactome.org/ReactomeGWT/entrypoint.html	[24]	Database of human pathways and reactions
GeneSNPs	http://www.genome.utah.edu/genesnps/		This Environmental Genome Project web resource integrates gene, sequence, and polymorphism data into individually annotated gene models. The human genes included are related to DNA repair, cell cycle control, cell signaling, cell division, homeostasis, and metabolism
Mouse Mutation Database	http://pathcuric1.swmed.edu/research/research.htm	[25]	The Database of mouse strains carrying targeted mutations in genes affecting cellular responses to DNA damage
BioCyc (EcoCyc, MetaCyc)	http://biocyc.org/	[26]	Experimentally studied metabolic pathways and enzymes from more than 1,500 organisms
BRENDA	http://www.brenda-enzymes.org/	[27]	The main collection of enzyme functional data
Pathway Commons	http://www.pathwaycommons.org/pc/	[28]	A collection of publicly available pathway data from multiple organisms
NGSethDB	http://bioinfo2.ugr.es/NGSmethDB/gbrowse/hg19/	[29]	Database for next-generation sequencing single-cytosine-resolution DNA methylation data
DNAreplication	http://DNAreplication.net/	[30]	Database for the eukaryotic DNA replication community
MethyCancer	http://methycancer.psych.ac.cn/	[31]	Links between DNA methylation levels and cancer
PubMeth	http://matrix.ugent.be/pubmeth/	[32]	Links between DNA methylation levels and cancer
MethDB (2009)	http://www.methdb.de/	[33]	The database for DNA methylation and environmental epigenetic effects
OriDB	http://www.oridb.org/index.php	[34]	Confirmed and predicted DNA replication origin sites
REBASE	http://rebase.neb.com/rebase/rebase.html	[35]	Enzymes and genes for DNA restriction and modification in prokaryotes
ROSPath	http://rospath.ewha.ac.kr/	[36]	Reactive oxygen species (ROS) signaling pathway proteins
Pathguide	http://www.pathguide.org/	[37]	A listing of pathway, signal transduction, and protein-protein interaction databases
CREMOFAC	http://www.jncasr.ac.in/cremofac/	[38]	Chromatin remodeling factors
DAnCER	http://wodaklab.org/dancer/	[39]	Disease-Annotated Chromatin Epigenetics Resource
Telomerase database	http://telomerase.asu.edu/	[40]	Sequences and structures of the RNA and protein subunits of telomerase, mutations of telomerase components
Replication Domain	http://www.replicationdomain.com/	[41]	Replication timing database and genome-wide data visualization tool

database (proteins, diseases, pathway steps, damage, etc.) are cross-referenced to the supporting literature and their respective primary databases. REPAIRtoire can be queried by the names of pathway, protein, enzymatic complex, damage and disease. The query tool returns a structured list of entries in the database that contain the query (e.g., "cancer", "DNA polymerase", "crosslink", "adenine", etc., or a name of the author).

The REPAIRtoire website provides a system for editing, adding, and removing data. These features have been provided for collaborators and "superusers" who are interested not only in viewing, but also curating the content of the database. Creating an account and logging into the database grants access to the administrative site of the database to a user. By entering the administration site, it is possible to add new data, delete information, edit, and correct mistakes. Editing information about proteins, genes, diseases, and types of damage is also available via wiki-like pages for particular database entries. Users can also add comments and suggest new references for the existing records. REPAIRtoire is unique in that it focuses on DNA repair and provides reciprocal annotation between damage entities and the proteins that can detect and remove them. It also contains more connections between DNA lesions and the respective proteins that can detect and remove them than can be found in general-purpose databases.

The REPAIRtoire website which also provides an online tool for drawing images of DNA-protein complexes (accessible via the "draw a picture" link in the main menu) is provided. This tool has been developed to illustrate all steps of DNA repair pathways as protein-DNA complexes, in which proteins are displayed in the textbook-like format of "potato models" (ellipsoids). However, it can be also used outside the DNA repair context to create images of any protein-protein or protein-DNA complexes. The drawing engine uses the SVG format provided by the W3C consortium and enables exporting the image in the JPEG format. Images created in the SVG vector format can be scaled without losing quality and can be modified with external tools for vector graphics processing, for example, Inkscape or other free or commercially available software.

The "*repairGENES*" database (http://www.repairgenes.org/) collects information about genes encoding proteins involved in DNA repair and connects information taken from sequence and ontology databases. At the moment, the site contains DNA repair genes from 134 selected species. The database can be browsed by organisms and by biological processes defined by the Gene Ontology (GO) standard [42]. The species are organized in a taxonomy tree. For processes, 17 subcategories of the GO term "DNA repair" (GO:0006281) and their respective subterms are distinguished. For each process, the organisms and genes that refer to this term can be listed. Also, it is possible to highlight the processes for a given organism. The major advantage of using GO terms is that they are being used ubiquitously for annotating sequence data. The raw data about DNA repair genes is extracted from the SWISS-PROT database. The repairGenes database also gives an overview of DNA repair processes and genes in five selected organisms (*Archaeoglobus fulgidus*, *Drosophila melanogaster*, *E. coli*, *Homo sapiens*, and *S. cerevisiae*), in total listing 452 genes.

"*Human DNA Repair Genes*" is an online supplement to a review published by Wood et al. in 2005 [17] and updated regularly (http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html). It provides a table with Gene Name (synonyms) linked to the GeneCards Human Gene Database at Cancer Research UK (http://bioinformatics.cancerresearchuk.org/genecards/) [43], activity linked to the OMIM database, chromosome location linked to the NCBI MapView, and an accession number linked to the NCBI Entrez server [44].

The "*Repair-FunMap*" database [23] used to provide information about the network of interactions between proteins involved in DNA repair and other proteins, but to our best knowledge it is no longer available.

2.2. General-Purpose Databases Relevant to DNA Repair. "*KEGG*" (Kyoto Encyclopedia of Genes and Genomes, available at http://www.genome.jp/kegg/) [21] is a collection of separate cross-linked databases including KEGG PATHWAY, KEGG DISEASE (human diseases), KEGG GENES (genes and proteins), and KEGG ORGANISMS. Of particular relevance to DNA repair are KEGG GENES (a catalog of genes for sequenced genomes obtained from publicly available resources, mostly NCBI RefSeq and KEGG PATHWAY (a collection of manually drawn pathway maps representing knowledge on the molecular interaction and reaction networks for: global map of pathways, metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and interaction of these systems with drugs)). DNA repair pathways annotated in KEGG include BER, NER, NHEJ, MMR, and HRR but not DDS, DDR, or TLS. A schematic graphical representation of protein-DNA complexes in the reaction steps of each pathway is available for Eukaryotes and Prokaryotes separately. KEGG BRITE is another component of KEGG that is important for analyzing DNA repair systems. It is a collection of hierarchical classifications representing knowledge on various aspects of biological systems. In contrast to KEGG PATHWAY, which is limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of relationships. The most relevant and interesting part is a section devoted to DNA repair (identifier "ko03400"—"DNA repair and recombination proteins"), where all DNA repair proteins available in KEGG are classified according to their functions in this process.

"*Reactome*" (http://www.reactome.org/ReactomeGWT/entrypoint.html) [24] is a resource developed in collaboration among different groups as an open source curated bioinformatics database of human pathways and reactions. The site provides bioinformatics tools for pathway analysis such as: the Pathway Browser, the Pathway and Expression Analysis tools, or the Species Comparison tool. In contrast to KEGG, Reactome includes graphical representations of DNA repair pathways generated for each organism explicitly instead of a generalized view like in KEGG. Moreover, Reactome divides pathways into subpathways, for example,

GG-NER (global genomic NER) in human is divided into four subpathways (DNA damage recognition, formation of the incision complex, dual incision reaction, and gap-filling DNA repair synthesis and ligation). Each subpathway contains individual reactions visualized in the context of the entire cellular metabolic map. The Pathway Analysis tool facilitates analysis of different pathways, for example, finding connections between RNA transcription and DNA repair, facilitating interdisciplinary studies [45].

The "*GeneSNPs*" database (http://www.genome.utah.edu/genesnps/) is dedicated to known human polymorphisms and has a section devoted to DNA repair. It can be accessed from the main page by selecting "DNA repair" in the Gene Lists menu on top of the home page. The SNP loci are presented as a table of 119 human genes involved in DNA repair and connected to phenotypes described in the OMIM database [44]. An exemplary usage of this resource is the study of polymorphisms in the DNA repair gene XRCC, where all the SNP data were collected from the GeneSNPs database [46]. More phenotypes of DNA repair defects can be found in the "*Mouse Mutation Database*" (a database of mouse strains carrying targeted mutations in genes affecting cellular responses to DNA damage available at http://pathcuric1.swmed.edu/research/research.htm) [25].

"*BioCyc*" (http://biocyc.org/) [26] is a collection of 1004 (as of February 2011) Pathway/Genome Databases. Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism. This is not only a collection of databases but of tools for bioinformatics analysis, including the following: a genome browser, a display of individual metabolic pathways and of full metabolic maps, visual analysis of user-supplied "omics" datasets by painting onto metabolic, regulatory, and genome maps, and comparative analysis tools. There is also downloadable version of BioCyc that includes the Pathway Tools. The BioCyc databases are divided into three tiers, based on their quality. Tier 1 databases have received person-decades of literature-based curation and are the most accurate. These include for example, EcoCyc (http://ecocyc.org/) [47], a comprehensive database of *Escherichia coli* K-12 MG1655 biology or MetaCyc (http://metacyc.org/), a database of nonredundant, experimentally elucidated metabolic pathways. Data included in these databases undergo a curation procedure involving external experts, who work on particular cellular systems to provide a comprehensive literature overview and up-to-date coverage of the field. Recently, this type of curation has been applied to the process of DNA repair; both direct repair mechanisms, such as photolyase, as well as indirect repair mechanisms, such as nucleotide excision repair, base excision repair and homologous recombination have been annotated [47]. Tier 2 and Tier 3 databases of BioCyc contain computationally predicted metabolic pathways, predictions as to which genes code for missing enzymes in metabolic pathways, and predicted operons. BioCyc does not include a dedicated DNA repair section, but information on DNA repair pathways can be found in other database sections. Data available in BioCyc can be used in in-depth analyses of biological systems relevant to different fields of research. This approach has been demonstrated in the study of differential network expression during drug and stress response by Cabusora et al. [48], where the expression data of known stress responders and DNA repair genes in mycobacterium tuberculosis from BioCyc collection were used.

"*BRENDA*" (BRaunschweig ENzyme Database, http://www.brenda-enzymes.org/) [27] is a comprehensive database on enzymes that collects manually annotated information on properties of enzymes, including mutants and engineered variants. It describes enzymes involved in DNA repair that have an E.C. number (e.g., uvrA: EC 3.1.25.1). Enzyme records contain data taken from the primary literature, such as classification, nomenclature, reaction type, substrate specificity, functional parameters, species, protein sequence and structure, practical application, information on mutants and engineered variants, stability, disease, isolation, and preparation. An essential part of BRENDA consists of information on metabolites and small molecules, which interact with enzymes as substrates and products, inhibitors, activating compounds, cofactors, or bound metals. BRENDA provides also enzyme disease-related information obtained from PubMed entries by text-mining procedures. BRENDA is currently the largest continuously maintained and publicly available enzyme database and covers a large number of experimentally characterized DNA repair enzymes.

"*Pathway Commons*" (http://www.pathwaycommons.org/pc/) is a comprehensive collection of publicly available pathway data from multiple organisms [28], which includes biochemical reactions, complex assembly, transport, catalysis events, and physical interactions involving proteins, DNA, RNA, small molecules, and complexes. This meta-database collects information from other databases such as Reactome or BioGrid, thereby facilitating analyses of system-level datasets across several species. It allows users to browse and search pathways across multiple valuable public pathway databases and download an integrated set of pathways in the BioPAX format for global analysis. It also provides an interface for software developers to create software for more advanced analyses and hence may be a very useful resource for programmatic linking of data on DNA repair systems with other cellular systems and pathways.

There exist numerous databases dedicated to other aspects of DNA metabolism. Examples include DNA replication (OriDB [34], ReplicationDomain [41]), apoptosis (Deathbase [49]), telomere maintenance (Telomerase database [40]), DNA restriction and modification (REBASE [35]), and epigenetics/chromatin modification (DAnCER [39]). These processes are relevant to DNA repair as they may contribute to DNA damage (replication) or regulation of other enzymatic processes (DNA methylation, cell cycle control, and apoptosis).

3. Bioinformatics Tools for the Study of DNA Repair Proteins

In addition to databases that store and disseminate the data, there are also bioinformatics tools that can be particularly useful for data analyses. We would like to emphasize three groups of predictive tools that can be particularly useful

for analyzing DNA repair enzymes: methods for predicting and modeling protein structures, predicting protein-DNA interactions and complexes, predicting the effect of amino acid substitutions on protein stability and function, and their phenotypic effect [50], as well as predicting cancer outcome [39].

3.1. Protein Structure Prediction. There is a large number of tools, with which to predict the structure of a protein when only its sequence is known. Their performance is evaluated in the biannual CASP benchmarking experiment [51]. One approach we would like to highlight here is homology modeling. There, a protein with known 3D structure is used as a template to construct a model for another, evolutionarily-related protein (a target). This approach requires not only an experimentally solved structure of the template protein, but also a pairwise sequence alignment between the target and the template. Among the numerous methods, the "*SWISS-MODEL*" server (http://swissmodel.expasy.org/) supports not only the fully automatic construction of homology models via its web interface, it also helps finding a suitable template and alignment [52]. It is particularly useful for building models of proteins that are closely related to the experimentally determined structures, so the relationship can be detected by methods such as "*BLAST*" [53]. If no such closely related templates are available, advanced template search and alignment tools such as "*HHSEARCH*" [54] can be used to identify remote evolutionary relationships. There are also specialized "meta-servers" such as the "*GeneSilico Metaserver*" [55] developed in the laboratory of the authors of this paper. These tools use several third-party methods and infer a consensus prediction.

As an example of protein modeling application to the analysis of DNA repair, we may refer to an analysis carried out in our laboratory: Missense alterations of the mismatch repair gene MLH1 have been identified in a significant proportion of individuals suspected of having Lynch syndrome, a hereditary syndrome that predisposes for cancer of colon and endometrium. The pathogenicity of many of these alterations was, however, unclear. A number of MLH1 alterations are located in the C-terminal domain (CTD) of MLH1, which is responsible for constitutive dimerization with another protein PMS2. We used the aforementioned "*GeneSilico Metaserver*" [55] to identified structurally characterized homologs of MLH1 and align their sequences, thereby enabling the construction of a homology model for MLH1 using the "FRankenstein's Monster" approach [56, 57]. That structural model was used to analyze 19 alterations connected to Lynch syndrome and to identify three alterations that decrease the efficiency of MMR in human by interfering with the MLH1-PMS2 dimerization, confirming that they are pathogenic, and suggesting that defective dimerization underlies their deleterious effect [50].

3.2. Methods for Predicting Protein-DNA Interactions. When analyzing enzymes acting on DNA, it is often important to know which parts of them interact with the substrate. Prediction of DNA-binding residues is facilitated by the knowledge of protein structure, either from experiment or from prediction (see above). An example of a bioinformatics online tool for structure-based prediction of DNA-binding residues is "*DISPLAR*" (http://pipe.scs.fsu.edu/displar.html), which uses a machine learning approach [58]. There are also methods, available as web services, enabling prediction of DNA-binding from protein sequence alone. Examples include "*BindN+*" (http://bioinfo.ggc.org/bindn+/) [59], "*DISIS*" (http://www.predictprotein.org/) [60], and "*DNABindR*" (http://turing.cs.iastate.edu/PredDNA/predict.html) [61].

If 3D structures of the components are known, it is also possible to obtain a three-dimensional model of protein-DNA complexes. The "*HADDOCK*" server (http://haddock.chem.uu.nl/) uses a flexible docking approach to build a complex from two or more separate protein and DNA structures [62]. It takes into account additional information such as distances between interacting residues and includes them as "ambiguous interaction restraints". This allows to use results from experimental analyses like mutation, crosslinking, and footprinting experiments or computational predictions made, for example, by the above-mentioned bioinformatics methods. It is important to note that HADDOCK generates a complex structure for all given components, but it does not evaluate whether the given components really interact and does not enable the modeling of large conformational changes. Also, identifying the correct interaction region is the most error-prone step, which is why accurate experimental knowledge is essential to obtain reliable structures. The HADDOCK developers also provide an extensive dataset of protein-DNA complexes that can be used for benchmarking purposes [63]. An alternative approach is to build models with other methods, without the use of experimental data, and then use the "*FILTREST3D*" method developed in the laboratory of the authors [64] to rank them according to the extent of agreement with the restraints.

3.3. Methods for Predicting the Effects of Amino Acid Substitutions. As illustrated by the example of the MLH1 protein, prediction of mutation/substitution effects on protein structure and function, and linking them to the relevant phenotype can be very useful in the study of DNA repair proteins. "*SNPs3D*" (http://www.snps3d.org/) [65] is an online tool that returns predictions of functional effects of nonsynonymous SNPs stored in the NCBI dbSNP database; currently it does not make predictions for altered sequences submitted by the users. There are a few predictive online methods that use protein structure (solved experimentally or modeled) to infer the effect of user-defined amino acid substitutions. "*CUPSAT*" (http://cupsat.tu-bs.de/) (Cologne University Protein Stability Analysis Tool) [66] predicts Gibbs-free energy changes associated with amino acid substitutions, based on analyzing of residue interactions with its 3D environment. "*PopMusic*" (http://babylone.ulb.ac.be/popmusic/) [67] evaluates the changes of protein stability resulting from single-residue or multiple substitutions. "*I-Mutant 2.0*" (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi) [68] also predicts protein stability changes upon single-site substitutions. It can be used both as a

classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related Gibbs-free energy changes. "*MUpro*" (http://www.ics.uci.edu/~baldig/mutation.html) [69] is a set of machine learning programs to predict how single-site amino acid substitutions affect protein stability. The server accepts single protein sequences or sequences with a predicted tertiary structure of the protein as an input.

There are also methods that predict mutation/substitution effects based on sequence information alone. "*PolyPhen*" (Polymorphism Phenotyping) (http://genetics.bwh.harvard.edu/pph/) [70] predicts the possible impact of an amino acid substitution on the structure and function of human proteins, based on straightforward empirical rules. "*SIFT*" (http://sift.jcvi.org/) [71] predicts whether an amino acid substitution (AAS) affects protein function based on analysis of sequence profiles. It can be applied to study naturally occurring nonsynonymous polymorphisms as well as laboratory-induced missense mutations. "*MutPred*" (http://mutpred.mutdb.org/) [72] is a web application that predicts the gain/loss of 14 different structural and functional properties (for instance, gain of helical propensity or loss of a phosphorylation site). It also classifies an amino acid substitution as disease-associated or neutral in human. "*PhD-SNP*" (http://snps.uib.es/phd-snp/PhD-SNP.html) [73] is another machine-learning method for predicting whether a phenotype derived from a nonsynonymous SNP could be related to a genetic disease in humans. It is optimized to predict if a given point mutation can be classified as a disease-related or a neutral polymorphism.

3.4. Predicting Cancer Outcome. A tool which facilitates the analyses of cancer-related proteins, genes and pathways is CAERUS [39]—a tool for predicting cancer outcomes using relationships between protein structural information, protein networks, gene expression data, and mutation data (http://www.oicr.on.ca/research/ouellette/caerus/). This tool was developed in order to identify a list of gene signatures and to better predict cancer by investigating the changes in gene expression profiles caused by disruptions between protein-protein interactions and domain-domain interactions in the human interactome. As the authors of CAERUS indicate, it was tested on a set of well-documented breast cancer patients, which suggests that the disrupted interactome is important to determine patient prognosis. They also declare that this approach is robust if tested on other independent data sets and therefore offers a promising prognostic tool to classify different cancer outcomes. As DNA repair is closely connected to cancer, this service can be used in the analysis of proteins and genes related to oncogenesis.

4. Summary

DNA repair is currently covered by a few dedicated databases. While REPAIRtoire and repairGenes focus on this topic, information is also available via general-purpose pathway databases. The main bottlenecks are the data collection and standardization. For instance, there is no specialized, universal ontology and no standards to describe entities and processes involved in DNA repair. Connecting the known "parts" such as enzymes, to pathways and processes in a formalized way that at the same time provides more insight into DNA repair processes, is probably the biggest challenge for the bioinformatics of DNA repair in the nearest future. It may be necessary to extend the currently established GO ontology by a vocabulary that will allow for describing repair processes on the protein complex and reaction level. A particular challenge is to find a consistent and appealing way to represent repair processes visually, and to include not only 3D descriptions, but also the dimension of time. The development and application of new computer programs for simulating and visualizing molecular processes involving multiple components will certainly contribute to our understanding of the complex process of DNA repair. In particular, it may help in the identification of new biomarkers, in predicting the possible side-effects of drugs based on personal genome information, and in the development of new therapeutic agents to restore the proper function of DNA repair proteins affected by disease-causing mutations.

Acknowledgments

The authors would like to thank Joanna Krwawicz for sharing her knowledge on DNA repair pathways and resources and for her invaluable contribution to the REPAIRtoire database project. K. Milanowska has been supported by the Foundation for Polish Science (Grant TEAM/2009-4/2). K. Rother has been supported by the German Academic Exchange Service (Grant D/09/42768). J. M. Bujnicki has been supported by the 7FP Grant "HEALTH-PROT" from the European Commission (contract number 229676).

References

[1] P. Jeggo and M. F. Lavin, "Cellular radiosensitivity: how much better do we understand it?" *International Journal of Radiation Biology*, vol. 85, no. 12, pp. 1061–1081, 2009.

[2] L. Maddukuri, D. Dudzińska, and B. Tudek, "Bacterial DNA repair genes and their eukaryotic homologues: 4. The role of nucleotide excision DNA repair (NER) system in mammalian cells," *Acta Biochimica Polonica*, vol. 54, no. 3, pp. 469–482, 2007.

[3] K. D. Arczewska and J. T. Kuśmierek, "Bacterial DNA repair genes and their eukaryotic homologues: 2. Role of bacterial mutator gene homologues in human disease. Overview of nucleotide pool sanitization and mismatch repair systems," *Acta Biochimica Polonica*, vol. 54, no. 3, pp. 435–457, 2007.

[4] N. C. Brissett and A. J. Doherty, "Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway," *Biochemical Society Transactions*, vol. 37, no. 3, pp. 539–545, 2009.

[5] A. Vaisman, A. R. Lehmann, and R. Woodgate, "DNA polymerases η and ι," *Advances in Protein Chemistry*, vol. 69, pp. 205–228, 2004.

[6] A. B. Robertson, A. Klungland, T. Rognes, and I. Leiros, "DNA repair in mammalian cells: base excision repair: the long and short of it," *Cellular and Molecular Life Sciences*, vol. 66, no. 6, pp. 981–993, 2009.

[7] J. Krwawicz, K. D. Arczewska, E. Speina, A. Maciejewska, and E. Grzesiuk, "Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease," *Acta Biochimica Polonica*, vol. 54, no. 3, pp. 413–434, 2007.

[8] R. Olinski, A. Siomek, R. Rozalski et al., "Oxidative damage to DNA and antioxidant status in aging and age-related diseases," *Acta Biochimica Polonica*, vol. 54, no. 1, pp. 11–26, 2007.

[9] B. Tudek, "Base excision repair modulation as a risk factor for human cancers," *Molecular Aspects of Medicine*, vol. 28, no. 3-4, pp. 258–275, 2007.

[10] R. de Bont and N. van Larebeke, "Endogenous DNA damage in humans: a review of quantitative data," *Mutagenesis*, vol. 19, no. 3, pp. 169–185, 2004.

[11] F. Drabløs, E. Feyzi, P. A. Aas et al., "Alkylation damage in DNA and RNA—repair mechanisms and medical significance," *DNA Repair*, vol. 3, no. 11, pp. 1389–1407, 2004.

[12] T. Lindahl, "Instability and decay of the primary structure of DNA," *Nature*, vol. 362, no. 6422, pp. 709–715, 1993.

[13] E. C. Friedberg et al., "DNA repair and mutagenesis," 2006.

[14] W. K. Hansen and M. R. Kelley, "Review of mammalian DNA repair and translational implications," *Journal of Pharmacology and Experimental Therapeutics*, vol. 295, no. 1, pp. 1–9, 2000.

[15] S. Raptis and B. Bapat, "Genetic instability in human tumors," *EXS*, no. 96, pp. 303–320, 2006.

[16] D. M. Wilson and D. Barsky, "The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA," *Mutation Research*, vol. 485, no. 4, pp. 283–307, 2001.

[17] R. D. Wood, M. Mitchell, and T. Lindahl, "Human DNA repair genes, 2005," *Mutation Research*, vol. 577, no. 1-2, pp. 275–283, 2005.

[18] R. D. Wood, M. Mitchell, J. Sgouros, and T. Lindahl, "Human DNA repair genes," *Science*, vol. 291, no. 5507, pp. 1284–1289, 2001.

[19] F. Chen, W. Q. Liu, A. Eisenstark, R. N. Johnston, G. R. Liu, and S. L. Liu, "Multiple genetic switches spontaneously modulating bacterial mutability," *BMC Evolutionary Biology*, vol. 10, no. 1, article 277, 2010.

[20] R. D. Wood, M. Mitchell, and T. Lindahl, "Human DNA repair genes," 2010.

[21] M. Kanehisa, M. Araki, S. Goto et al., "KEGG for linking genomes to life and the environment," *Nucleic Acids Research*, vol. 36, no. 1, pp. D480–D484, 2008.

[22] K. Milanowska, J. Krwawicz, G. Papaj et al., "REPAIRtoire—a database of DNA repair pathways," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D788–D792, 2011.

[23] L. Wen and J. A. Feng, "Repair-FunMap: a functional database of proteins of the DNA repair systems," *Bioinformatics*, vol. 20, no. 13, pp. 2135–2137, 2004.

[24] L. Matthews, G. Gopinath, M. Gillespie et al., "Reactome knowledgebase of human biological pathways and processes," *Nucleic Acids Research*, vol. 37, no. 1, pp. D619–D622, 2009.

[25] E. C. Friedberg and L. B. Meira, "Database of mouse strains carrying targeted mutations in genes affecting biological responses to DNA damage Version 7," *DNA Repair*, vol. 5, no. 2, pp. 189–209, 2006.

[26] R. Caspi, T. Altman, J. M. Dale et al., "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases," *Nucleic Acids Research*, vol. 38, no. 1, Article ID gkp875, pp. D473–D479, 2009.

[27] M. Scheer, A. Grote, A. Chang et al., "BRENDA, the enzyme information system in 2011," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D670–D676, 2011.

[28] E. G. Cerami, B. E. Gross, E. Demir et al., "Pathway Commons, a web resource for biological pathway data," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D685–D690, 2011.

[29] M. Hackenberg, G. Barturen, and J. L. Oliver, "NGSmethDB: a database for next-generation sequencing single-cytosine-resolution DNAmethylation data," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D75–D79, 2011.

[30] S. Cotterill and S. E. Kearsey, "DNAReplication: a database of information and resources for the eukaryotic DNA replication community," *Nucleic Acids Research*, vol. 37, no. 1, pp. D837–D839, 2009.

[31] X. He, S. Chang, J. Zhang et al., "MethyCancer: the database of human DNA methylation and cancer," *Nucleic Acids Research*, vol. 36, no. 1, pp. D836–D841, 2008.

[32] M. Ongenaert, L. Van Neste, T. de Meyer, G. Menschaert, S. Bekaert, and W. van Criekinge, "PubMeth: a cancer methylation database combining text-mining and expert annotation," *Nucleic Acids Research*, vol. 36, no. 1, pp. D842–D846, 2008.

[33] C. Grunau, E. Renault, A. Rosenthal, and G. Roizes, "MethDB—a public database for DNA methylation data," *Nucleic Acids Research*, vol. 29, no. 1, pp. 270–274, 2001.

[34] C. A. Nieduszynski, S. I. Hiraga, P. Ak, C. J. Benham, and A. D. Donaldson, "OriDB: a DNA replication origin database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D40–D46, 2007.

[35] R. J. Roberts, T. Vincze, J. Posfai, and D. Macelis, "REBASE-A database for DNA restriction and modification: enzymes, genes and genomes," *Nucleic Acids Research*, vol. 38, no. 1, Article ID gkp874, pp. D234–D236, 2009.

[36] E. Paek, J. Park, and K. J. Lee, "Multi-layered representation for cell signaling pathways," *Molecular and Cellular Proteomics*, vol. 3, no. 10, pp. 1009–1022, 2004.

[37] G. D. Bader, M. P. Cary, and C. Sander, "Pathguide: a pathway resource list," *Nucleic Acids Research*, vol. 34, pp. D504–D506, 2006.

[38] A. Shipra, K. Chetan, and M. R. S. Rao, "CREMOFAC—a database of chromatin remodeling factors," *Bioinformatics*, vol. 22, no. 23, pp. 2940–2944, 2006.

[39] A. L. Turinsky, B. Turner, R. C. Borja et al., "DAnCER: disease-annotated chromatin epigenetics resource," *Nucleic Acids Research*, vol. 39, supplement 1, pp. D889–D894, 2011.

[40] J. D. Podlevsky, C. J. Bley, R. V. Omana, X. Qi, and J. L. Chen, "The Telomerase Database," *Nucleic Acids Research*, vol. 36, no. 1, pp. D339–D343, 2008.

[41] N. Weddington, A. Stuy, I. Hiratani, T. Ryba, T. Yokochi, and D. M. Gilbert, "ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data," *BMC bioinformatics*, vol. 9, p. 530, 2008.

[42] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[43] M. Safran, I. Dalah, J. Alexander et al., "GeneCards version 3: the human gene integrator," *Database: The Journal of Biological Databases and Curation*, vol. 2010, Article ID baq020, 2010.

[44] E. W. Sayers, T. Barrett, D. A. Benson et al., "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 37, no. 1, pp. D5–D15, 2009.

[45] L. D. Stein, "Using the Reactome database," *Current Protocols in Bioinformatics*, chapter 8, p. Unit 8.7, 2004.

[46] W. Ladiges, J. Wiley, and A. MacAuley, "Polymorphisms in the DNA repair gene XRCC1 and age-related disease," *Mechanisms of Ageing and Development*, vol. 124, no. 1, pp. 27–32, 2003.

[47] I. M. Keseler, C. Bonavides-Martínez, J. Collado-Vides et al., "EcoCyc: a comprehensive view of Escherichia coli biology," *Nucleic Acids Research*, vol. 37, no. 1, pp. D464–D470, 2009.

[48] L. Cabusora, E. Sutton, A. Fulmer, and C. V. Forst, "Differential network expression during drug and stress response," *Bioinformatics*, vol. 21, no. 12, pp. 2898–2905, 2005.

[49] J. Díez, D. Walter, C. Mũoz-Pinedo, and T. Gabaldón, "Editorial: DeathBase: a database on structure, evolution and function of proteins involved in apoptosis and other forms of cell death," *Cell Death and Differentiation*, vol. 17, no. 5, pp. 735–736, 2010.

[50] J. Kosinski, I. Hinrichsen, J. M. Bujnicki, P. Friedhoff, and G. Plotz, "Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair," *Human Mutation*, vol. 31, no. 8, pp. 975–982, 2010.

[51] S. Shi et al., "Analysis of CASP8 targets, predictions and assessment methods," *Database*, vol. 2009, Article ID bap003, 2009.

[52] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, no. 2, pp. 195–201, 2006.

[53] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[54] J. Söding, "Protein homology detection by HMM-HMM comparison," *Bioinformatics*, vol. 21, no. 7, pp. 951–960, 2005.

[55] M. A. Kurowski and J. M. Bujnicki, "GeneSilico protein structure prediction meta-server," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3305–3307, 2003.

[56] J. Kosinski, I. A. Cymerman, M. Feder, M. A. Kurowski, J. M. Sasin, and J. M. Bujnicki, "A "FRankenstein's Monster" approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation," *Proteins: Structure, Function and Genetics*, vol. 53, no. 6, pp. 369–379, 2003.

[57] J. Kosinski, M. J. Gajda, I. A. Cymerman et al., "FRankenstein becomes a cyborg: the automatic recombination and realignment of fold recognition models in CASP6," *Proteins: Structure, Function and Genetics*, vol. 61, no. 7, pp. 106–113, 2005.

[58] Y. Xiong, J. Liu, and D. Q. Wei, "An accurate feature-based method for identifying DNA-binding residues on protein surfaces," *Proteins: Structure, Function and Bioformatics*, vol. 79, no. 2, pp. 509–517, 2011.

[59] L. Wang, C. Huang, M. Q. Yang, and J. Y. Yang, "BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features," *BMC Systems Biology*, vol. 4, no. 1, article S3, 2010.

[60] Y. Ofran, V. Mysore, and B. Rost, "Prediction of DNA-binding residues from sequence," *Bioinformatics*, vol. 23, no. 13, pp. i347–i353, 2007.

[61] C. Yan, M. Terribilini, F. Wu, R. L. Jernigan, D. Dobbs, and V. Honavar, "Predicting DNA-binding sites of proteins from amino acid sequence," *BMC Bioinformatics*, vol. 7, article 262, 2006.

[62] S. J. de Vries, M. van Dijk, and A. M. Bonvin, "The HADDOCK web server for data-driven biomolecular docking," *Nature Protocols*, vol. 5, no. 5, pp. 883–897, 2010.

[63] M. van Dijk and A. M. J. J. Bonvin, "Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance," *Nucleic Acids Research*, vol. 38, no. 17, Article ID gkq222, pp. 5634–5647, 2010.

[64] M. J. Gajda, I. Tuszynska, M. Kaczor, A. Y. Bakulina, and J. M. Bujnicki, "FILTREST3D: discrimination of structural models using restraints from experimental data," *Bioinformatics*, vol. 26, no. 23, Article ID btq582, pp. 2986–2987, 2010.

[65] P. Yue, E. Melamud, and J. Moult, "SNPs3D: candidate gene and SNP selection for association studies," *BMC Bioinformatics*, vol. 7, article 166, 2006.

[66] V. Parthiban, M. M. Gromiha, and D. Schomburg, "CUPSAT: prediction of protein stability upon point mutations," *Nucleic Acids Research*, vol. 34, pp. W239–W242, 2006.

[67] D. Gilis and M. Rooman, "PoPMuSiC, an algorithm for predicting protein mutant stability changes. Application to prion proteins," *Protein Engineering*, vol. 13, no. 12, pp. 849–856, 2000.

[68] E. Capriotti, P. Fariselli, and R. Casadio, "I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure," *Nucleic Acids Research*, vol. 33, no. 2, pp. W306–W310, 2005.

[69] J. Cheng, A. Randall, and P. Baldi, "Prediction of protein stability changes for single-site mutations using support vector machines," *Proteins: Structure, Function and Genetics*, vol. 62, no. 4, pp. 1125–1132, 2006.

[70] I. A. Adzhubei, S. Schmidt, L. Peshkin et al., "A method and server for predicting damaging missense mutations," *Nature Methods*, vol. 7, no. 4, pp. 248–249, 2010.

[71] P. C. Ng and S. Henikoff, "Predicting deleterious amino acid substitutions," *Genome Research*, vol. 11, no. 5, pp. 863–874, 2001.

[72] B. Li, V. G. Krishnan, M. E. Mort et al., "Automated inference of molecular mechanisms of disease from amino acid substitutions," *Bioinformatics*, vol. 25, no. 21, pp. 2744–2750, 2009.

[73] E. Capriotti, R. Calabrese, and R. Casadio, "Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information," *Bioinformatics*, vol. 22, no. 22, pp. 2729–2734, 2006.